

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1401	100.0	266	2	088109	mycobacteri
2	1401	100.0	266	2	07BR87	07br87 mycobacteri
3	1401	100.0	266	2	Q7B2P1	Q7b2p1 mycobacteri
4	788	56.2	256	1	YT57MYCTU	Y57MYCTU
5	319	22.8	251	1	Q8A3K7	Q8a3k7 bacteroides
6	308	22.0	257	2	Q64MN8	Q64mn8 bacteroides
7	298.5	21.3	297	2	Q93Q25	Q93q25 salmonella
8	298	21.3	297	2	Q9S520	Q9s520 escherichia
9	294	21.0	293	2	Q72WK9	Q72wk9 desulfobact
10	287	20.5	248	2	Q8P241	Q8p241 methanosarc
11	276	19.7	265	2	Q8VXT0	Q8vxt0 leptospira
12	275	19.6	265	2	Q8VWTX	Q8vtx0 leptospira
13	275	19.6	265	2	Q8VXT7	Q8vtx7 leptospira
14	275	19.6	265	2	Q9AEE4	Q9aee4 leptospira
15	275	19.6	265	2	Q72QI4	Q72qi4 leptospira
16	274	19.6	265	2	Q8VWT9	Q8vwt9 leptospira
17	271	19.3	265	2	Q8VL41	Q8vl41 leptospira
18	267.5	19.1	248	2	Q9EXY1	Q9exy1 escherichia
19	267.5	19.1	258	2	Q7WYS0	Q7wys0 rhizobium
20	267.5	19.1	265	2	Q8VWT8	Q8vwt8 leptospira
21	266.5	19.0	248	2	Q9EXY4	Q9exy4 escherichia
22	257.5	18.4	248	2	Q7MW87	Q7mw87 porphyromon
23	255.5	18.2	251	2	Q9ZGK3	Q9zgz3 leptospira
24	255	18.2	254	2	Q9EXZ1	Q9exz1 salmonella
25	255	18.2	274	2	Q7NEU1	Q7neu1 gloeobacter
26	254.5	18.2	263	2	Q95AF9	Q95af9 leptospira
27	252	18.0	254	2	Q9EXY9	Q9exy9 salmonella
28	247	17.7	254	2	Q9XDJ3	Q9xdj3 bacteroides
29	247	17.6	276	2	Q9RQ94	Q9rq94 rhizobium
30	240	17.1	295	2	Q74BU3	Q74bu3 gloeobacter
31	238.5	17.0	247	2	Q66DN1	Q66dn1 versinia

```

RESULT 2
Q7BR87 ID Q7BR87 PRELIMINARY; PRT; 266 AA.
AC Q7BR87;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Glycosyltransferase gtfD.
OS Name=gtfD;
GN Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TMC 724, and 2151;
RX MEDLINE=22885473; PubMed=14523113; DOI=10.1099/mic.0.26528-0;
RA Eckstein T.M., Belisle J.T., Inamine J.M.;
RT "Proposed pathway for the biosynthesis of serovar-specific
RL glycopeptidolipids in Mycobacterium avium serovar 2.";
DR EMBL; AF125999; AAD20376.1; -.
DR EMBL; AF143772; AAD44222.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001173; Glyco_transf_2; 1.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e-118;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPVFSIIIPTFNAAVTLQACLGSIQVGYREVEVWLVGSGSTDTLTDIANSFRPELGS 60
Db 1 MTAPVFSIIIPTFNAAVTLQACLGSIQVGYREVEVWLVGSGSTDTLTDIANSFRPELGS 60
QY 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
Db 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
QY 121 GDVWRSTKSRHAGPDLRLFFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDENIR 180
Db 121 GDVWRSTKSRHAGPDLRLFFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDENIR 180
QY 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRMLAFLKD 240
Db 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRMLAFLKD 240
QY 241 KENRRALRTRLIIRKAVSKERSAEP 266
Db 241 KENRRALRTRLIIRKAVSKERSAEP 266

RESULT 3
Q7B2P1 ID Q7B2P1 PRELIMINARY; PRT; 266 AA.
AC Q7B2P1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Gsd protein (Hypothetical protein).
GN Name=gsd; OrderedLocusNames=MAP1234;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M. Bull T.;
RX Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
RA Hermon-Taylor J.;
RT "A low G C content element in Mycobacterium avium subsp.

```

```

RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
RT in M. tuberculosis.";
RL Microbiology 144:3413-3423 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Bull T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=k10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223833; CAAL1578.1; -.
DR EMBL; AE017231; AAS03551.1; -.
DR InterPro; IPR001173; Glyco_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e-118;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPVFSIIIPTFNAAVTLQACLGSIQVGYREVEVWLVGSGSTDTLTDIANSFRPELGS 60
Db 1 MTAPVFSIIIPTFNAAVTLQACLGSIQVGYREVEVWLVGSGSTDTLTDIANSFRPELGS 60
QY 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
Db 61 RLIVHSGPDDGPDYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVY 120
QY 121 GDVWRSTKSRHAGPDLRLFFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDENIR 180
Db 121 GDVWRSTKSRHAGPDLRLFFETNLCHQSIFYRRELFDGIGPYNLRVYRWADWDENIR 180
QY 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRMLAFLKD 240
Db 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRMLAFLKD 240
QY 241 KENRRALRTRLIIRKAVSKERSAEP 266
Db 241 KENRRALRTRLIIRKAVSKERSAEP 266

RESULT 4
YT57_MYCTU STANDARD; PRT; 256 AA.
ID YT57_MYCTU
AC Q50459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Putative glycosyl transferase Rv2957/MT3031/Mb2981 (EC 2.-.-.-).
GN OrderedLocusNames=Rv2957, MT3031, Mb2981;
ORFNames=MTCY349.31C, u0002kc;
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,

```

[illegible]


```
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OG Plasmid pDV.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.F., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough."
RL Nat. Biotechnol. 22:554-559(2004).
RL EMBL; AE017286; AAS94467.1; -.
DR TIGR; DVUA0080; -.
DR InterPro; IPR001173; Glyco trans 2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Complete proteome.
SQ SEQUENCE 293 AA; 31860 MW; 200AC62ECC830029 CRC64;

Query Match 21.0%; Score 294; DB 2; Length 293;
Best Local Similarity 33.6%; Pred. No. 3e-18;
Matches 89; Conservative 31; Mismatches 99; Indels 46; Gaps 8;

QY 1 MTAPVFSIIPTFNAAVTLOACLSIGVQTYREVEVLVDGGSTDRTLDIANSFRPELGS 60
Db 1 MNQPFSLVVVIAFENVAEVLRECLHSLASQTFRDFEVIQDGASTDGTTLAVAQEAAGRL-P 59

QY 61 RLNVHSGDDGPDYDAMNRGVGATGEVWFLFGADDTLYEPTTLAQVAFLGDHAAASHL-- 118
Db 60 RLUSLASVADSGIYDANRALPRVTDGWLFLGSDDTLAGPDVLAQCARRLADLPQVLYA 119

QY 119 -----VGDVVVRSTKSRHAGFPDLDRLLFETNLCHOSIFYRRELFDGIGPNLYRVW 172
Db 120 CGDALEVDPGSSRAVTCAGCA--TRMADHIPFCHSLWRACLFEA-RHFDATLIA 176

QY 173 ADWDFNIRCF-----SNPA--LITRYMDVWVISEYNDMTGFSNR 208
Db 177 ADYDLICRTWPHDGVGHGTGLITVTHMGVGVSSPAHRLRTQW-----ELVRVAAR 227

QY 209 QGTDFEPRKGL-PNFWAGWETCR 232
Db 228 HGGAFSRRRLVPLCKAVGLWVLCR 252

RESULT 10
QBP241
ID QBP241 PRELIMINARY; PRT; 248 AA.
AC QBP241;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Glycosyltransferase involved in cell wall biogenesis (EC 2.4.---
DE ).
GN OrderedLocusNames=MM0653;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Werkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierer A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
```

```
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea."
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR EMBL; AE013290; AAM30349.1; -.
DR InterPro; IPR001173; Glyco_trans 2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Complete proteome.
SQ SEQUENCE 248 AA; 29073 MW; 4C33553DBDA56D84 CRC64;

Query Match 20.5%; Score 287; DB 2; Length 248;
Best Local Similarity 28.6%; Pred. No. 1.1e-17;
Matches 65; Conservative 53; Mismatches 99; Indels 10; Gaps 5;

QY 4 PVFSIIPTFNAAVTLOACLSIGVQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 63
Db 3 PKVSIVITCYNAEFENAIYSLVNSYENIEYIIIDGASKDNTVSIINKYRSVMACPI- 61

QY 64 VHSQDDGPDYDAMNRGVGATGEVWFLFGADDTLYEPTTLAQVAFLGDHAAASHLYGDDV 123
Db 62 --SEPDKMGYEAMNKGKAATGDILYFLNSDDVFYDEVVENVVKVQKNNVELIYGPI 119

QY 124 VWR--STKSRHAGPDLDRLLFETN-LCHOSIFYRRELFDGIGPNLYRVWADDFNI 179
Db 120 IIRDTTNEFIKPHDYITKSYFIYNSICQQGIFYKASVFNKCGQFNDTYKIVGDYEWEL 179

QY 180 RCFNSPALITRYMDVWVISEYNDMTGFSMRQGTD---KEFRKRLPMYF 223
Db 180 RAFYKYNIRKRYEGIIAIFRDGGMCNCDKFSDLHAKERTKVIKEYF 226

RESULT 11
QBP241
ID QBP241 PRELIMINARY; PRT; 265 AA.
AC QBP241;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
DE Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hond Utrecht IV;
RX MEDLINE=21369626; PubMed=11476985; DOI=10.1016/S0928-8244(01)00245-0;
RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
RT "Genetic differences among the LPS biosynthetic loci of serovars of
RT Leptospira interrogans and Leptospira borgpetersenii."
RL FEMS Immunol. Med. Microbiol. 31:73-81(2001).
DR EMBL; AF316559; AAL49452.1; -.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 31424 MW; F85D0153E0197DDB CRC64;

Query Match 19.7%; Score 276; DB 2; Length 265;
Best Local Similarity 32.3%; Pred. No. 1.1e-16;
Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;

QY 4 PVFSIIPTFNAAVTLOACLSIGVQTYREVEVLVDGGSTD-----RTLDIANSFRP 56
Db 6 PKISITITLNNLEGRKLTLESVKSTQTYTNFELIVDGGSTDGSPEYLSNLDLIKFTS 65

QY 57 ELGSLRVHSGDDGPDYDAMNRGVGATGEVWFLFGADDTLYEPTTLAQVAFLGDHAAAS 116
Db 66 E-----KDKGIYNAQNKGISLSGSEYLVFLNAGDTLLQKNVLEISKFLQDDV-- 113

QY 117 HLVTGDDVVRSTKSRHAGFPDLDRLLF-----ETNLCHOSIFYRRELFDGIGPNLYRVW 172
Db 114 DLVTGNILDSKNGIIRKYPDRNLNPFYMWIKSLCHQAVFIRKSLFDLYGYNEEYLF 173

QY 173 ADWDFNIRCFNSPALITRYMDVWVISEYNDMTGFSM-----RQGTDKFRKRLPMYF 223
```


Search completed: February 27, 2005, 21:10:50
Job time : 177 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 21:01:28 ; Search time 40 Seconds
(without alignments)
639.841 Million cell updates/sec

Title: US-10-805-311-24

Perfect score: 1401

Sequence: 1 MTAPVFSIIPTFNAATLQ.....ALRTRLIRKAVSKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79:*

2: pir1:*

3: pir2:*

4: pir3:*

5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	788	56.2	247	2 B70670	hypothetical prote
2	238.5	17.0	275	2 AG0376	probable glycosyl
3	232.5	16.6	260	2 E71975	hypothetical prote
4	231.5	16.5	259	2 F64532	conserved hypothet
5	220	15.7	324	2 AB2190	hypothetical prote
6	207	14.8	321	2 AB2188	hypothetical prote
7	206.5	14.7	316	2 AE2189	hypothetical prote
8	205.5	14.7	298	2 B75096	glycosyl transfera
9	204.5	14.6	330	2 AH2188	hypothetical prote
10	202.5	14.5	262	2 E70714	hypothetical prote
11	196.5	14.0	337	2 AG1920	hypothetical prote
12	196.5	14.0	367	2 G95948	probable glycosyl
13	192	13.7	336	2 A97168	glycosyltransferas
14	190.5	13.6	322	2 AE2160	glycosyltransferas
15	190.5	13.6	323	2 AD2189	hypothetical prote
16	190	13.6	248	2 C90984	probable glycosyl
17	190	13.6	248	2 F85829	glycosyl transfera
18	189.5	13.5	333	2 AH2026	hypothetical prote
19	188.5	13.5	333	2 B97168	glycosyltransferas
20	188	13.4	299	2 B83557	probable glycosyl
21	182	13.0	344	2 AC0974	probable glycosyl
22	178	12.7	356	2 S74766	hypothetical prote
23	177.5	12.7	343	2 AI2091	glycosyltransferas
24	175	12.5	318	2 AG2189	hypothetical prote
25	174.5	12.5	333	2 G86651	sugar transferase
26	174.5	12.5	623	2 AH1209	B. subtilis minor
27	174.5	12.5	996	2 S76194	hypothetical prote
28	173.5	12.4	324	2 AE9290	probable glycosyl
29	173.5	12.4	338	2 E91190	probable regulator

ALIGNMENTS

RESULT 1

B70670 Hypothetical protein Rv2957 - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: u0002kc protein

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70670; S73064

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, A.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: B70670

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-275 <COL>

A: Cross-references: GB:Z83018; GB:AL123456; NID:g3261671; PIDN: CAB05419.1; PID: g1694875

A: Experimental source: strain H37RV

R: Smith, D.R.; Robison, K.

A: Submitted to the EMBL Data Library, September 1994

A: Description: Mycobacterium tuberculosis cosmid tbc2.

A: Reference number: S73053

A: Accession: S73064

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-275 <SMI>

A: Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50938.1; PID:g560517

C: Genetics:

A: Gene: Rv2957

C: Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 56.2%; Score 788; DB 2; Length 275;
Best Local Similarity 60.9%; Pred. No. 4.1e-65;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY	1	MTAPVFSIIPTFNAATLQACLSIVGQTVREVEVLVDGSGSTDRDLDIANSFRPELGS	60
DB	20	MAAPMFIIITPLNVAALPACLDSIARQTCGDFELVLVDGSGSTDETLDIANIFAPNLGE	79
QY	61	RLVHSGPDDGYDAMNNGVATGCVLFLGADDTLVEPTTLAQVAFLGHPAASHLVY	120
DB	80	RLIIHRDTQGYDMMNNGVDLATGTWLLFLGADDSLYEADTLARVAALFIGHEFSDLVY	139
QY	121	GDVNRSTKSRHAGPDLRLFFETNLCHQISFYRRELFDDGIGPYNLRVYRWADFNIR	180
DB	140	GDVIMRSTNFRWGGAFLDRLLEKNIHQALFYRGLFGTIGPYNLRVRLADFNIR	199
QY	181	CFSNPALITRMVDVVISEYNDMTGSMRQGTDKPRKRLPMYFVWAGWETCRMLAFLKD	240
DB	200	CFSNPALVTRYMHVVVASYNEFGGLS-NTIIVDKFLKRLPM-----	239

30	173.5	12.4	338	2	F86037	probable regulator
31	172.5	12.3	344	1	Q3ECTH	hypothetical 40.5k
32	172	12.3	573	2	E89789	hypothetical prote
33	171	12.2	344	2	A70037	capsular polysacch
34	170.5	12.2	301	2	F95205	glycosyl transfera
35	170	12.1	354	2	H96021	probable glycosyl
36	169.5	12.1	774	2	AC1566	hypothetical prote
37	168.5	12.0	294	2	G71148	probable glycosyl
38	168	12.0	250	2	A64099	hypothetical prote
39	168	12.0	318	2	AH2189	glycosyl transfera
40	168	12.0	462	2	AH1566	hypothetical prote
41	168	12.0	1013	2	AE1876	galactosamine-cont
42	165.5	11.8	340	2	T44330	hypothetical prote
43	165	11.8	303	2	A84114	glycosyl transfera
44	165	11.8	323	1	H64130	glycosyl transfera
45	164	11.7	334	1	G71153	hypothetical prote

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2190

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-324 <KUR>

A;Cross-references: UNIPROT:Q8YSL1; GB:BA000019; PIDN:BA074772.1; PID:g17132167; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3073

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 15.7%; Score 220; DB 2; Length 324;

Best Local Similarity 25.9%; Pred. No. 1.4e-12;

Matches 69; Conservative 53; Mismatches 108; Indels 36; Gaps 8;

QY 4 PVFSIIPTFNAAVTLQACLGSIYGVQTYREVEVLVDGSGTDRTLDIANSFRPELGSRLV 63

DB 2 PKTSVILPAYNAERTILETINSVLNQTFSDELIIVINDGSTDRTLEVLQVND---DARLK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLYVGDV 123

DB 59 VYSYNSRASCARNHGISHAVGDFISFLDADD-LWTFDKLEQLSALNNHPEAGVAYSWT 117

QY 124 VMRSTKSRHAGPFD-----LDRL--FETWLCHQSIFYRRELFDDGIGPNLYRVRWA 173

DB 118 YTIDDKGEILKPEPLVEGNNVYTDLLANFLTNGSNPLI--RKAALIASIGFDDTLRSGE 175

QY 174 DWDFNIRCFSPNALITRYMDVWVISEYNDMTGFSMRQGTDEKFRKRLPMYFWAGWETCRR 233

DB 176 DWYWLRL-----LAKYFVFWVQKHQIL---YRRSVTSKSKLQI-----IRE 215

QY 234 MIAFLDKENRRRLALRLIRKAVS 259

DB 216 ASLAILDKAMKVLPLELQVKKHSL 241

RESULT 6

AG2188

hypothetical protein alr3062 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: AG2188

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG2188

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-321 <KUR>

A;Cross-references: UNIPROT:Q8YSM2; GB:BA000019; PIDN:BA074761.1; PID:g17132156; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3062

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.8%; Score 207; DB 2; Length 321;

Best Local Similarity 32.0%; Pred. No. 2.2e-11;

Matches 66; Conservative 35; Mismatches 83; Indels 22; Gaps 7;

QY 4 PVFSIIPTFNAAVTLQACLGSIYGVQTYREVEVLVDGSGTDRTLDIANSFRPELGSRLV 63

DB 2 PKSVVIVPAYNAMPYLETLESVLRTYHDFEVLVNDGSSDNTTEWWSQI---LDPLRK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLYVGDV 123

DB 59 LISQANOGLAGARNTGIVNASGEYIAFLDADD-IWEPTKLAQVSLDENPTVGLVYTWV 117

QY 124 VMRSTKSRHAGPFDLRL-----LFETNL--CHQSIFYRRELFDDGIGPY--NLYRVRW 172

DB 118 AYIDEQKSTGKIFKNQVEGYVMPQLTEHNIVECGSVALVRRVCFKMGFLFORNLGSYY- 176

QY 173 ADWDFNIRCFSPNALITRYMDVWVISE 198

DB 177 EDMDWMLR-----IATSYDFKVKKE 196

RESULT 7

AE2189

hypothetical protein alr3068 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: AE2189

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE2189

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-316 <KUR>

A;Cross-references: UNIPROT:Q8YSL6; GB:BA000019; PIDN:BA074767.1; PID:g17132162; GSPDB:

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3068

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.7%; Score 206.5; DB 2; Length 316;

Best Local Similarity 29.6%; Pred. No. 2.4e-11;

Matches 55; Conservative 37; Mismatches 81; Indels 13; Gaps 5;

QY 4 PVFSIIPTFNAAVTLQACLGSIYGVQTYREVEVLVDGSGTDRTLDIANSFRPELGSRLV 63

DB 2 PTISVILPAYNAERTILETINSVQQQTFSDFLIINDGSTDRTLELQNTIRDE---RLK 58

QY 64 VHSPPDGPYDAMNRGVGATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLYVGDV 123

DB 59 IFSYENGGLCTARNRGISHASGEPIAFLDADDLWTHDKLEQLTA-LQHPPEAGVAYSWT 117

QY 124 VMRSTKSRHAGP-----FDLD---RLLETWLCHQS--IFYRRELFDDGIGPNLYRVRWAD 174

DB 118 YFMDSQGSSTPGVSLFFEGDVQAHLVNNFLASGNSPLIRKQAIYESVGFEFSDNCMGCAD 177

QY 175 WDFNIR 180

DB 178 WDYWLRL 183

RESULT 8

B75096

Glycosyl transferase PAB0772 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C;Accession: B75096

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str

A;Reference number: A75001

A;Accession: B75096

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-298 <KAW>

A;Cross-references: UNIPROT:Q9UZ16; GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB500

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB0772

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 14.7%; Score 205.5; DB 2; Length 298;

Best Local Similarity 29.0%; Pred. No. 2.8e-11;

Matches 71; Conservative 38; Mismatches 99; Indels 37; Gaps 10;

Qy	4	PVFSIIIPFNAAVTIQA	CLGSI	VGQTYRE	VVLVDGGSTDR	TDLTDANGSRPELGSRLV	63
Dd	5	PKVISYMPYNSSRYLA	EIESHLDQTYNDFE	FIIIVDDGSTDSYNTISSYANK-DNR	I	63	
Qy	64	VHSGPDPGPYDAMNR	GVGATGEWLF	FLGADD-----TLVEPTT	LAQVAALGHAA	115	
Dd	64	VISREHGLVDSLNEG	INIARGKVIARMDAD	DISINNRIEKQEFELNKDV	DLIGTRIE	123	
Qy	116	SHLVYGDUVMRST---	KSRHAGFPD	LRL--LPETN--LCHQSIF	YRRELFDGIGPNLR	168	
Dd	124	A---FGDI	EKKQTIYNSAFS	IKFSDSQIEQVELTSCAIPHSPVMFKOSIVKLURGYRKE	180		
Qy	169	YRWADWDENIRCFNS	PALITRMV	WVISEYN-----DMTSFMROG----	210		
Dd	181	YDTAEYDLWLRAIR	NGYKIVR-MDECLI	KYRVHNKSKTAVEMFNPKQVEYTMKAKIDYI	239		
Qy	211	TKQEPKRKLPMYFWA-	-GWETCRMLAPLKDKEN	243			
Dd	240	NDTNKKDKVDYLWG	ASTGGKLVKKVVEST	TDKFN	274		

RESULT 14
AE2160
glycosyltransferase alr-2836 [imported] - Noetoc sp. (strain PCC 7120)
C.Species: Noetoc sp. PCC 7120
A;Note: Noetoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE2160
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;
DNA Res. 8. 205-213. 2001

A;Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Nostoc punctiforme* strain PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE2160
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <KUR>
A;Cross-references: UNIPROT:P22639; GB:BA000019; PIDN:BA874535.1; PID:g171313930
A;Experimental source: strain PCC 7120
C;Genetics
A;Gene: alr2836

Query Match 13.6%; Score 190.5; DB 2; Length 322;
Best Local Similarity 29.1%; Pred. No. 7.5e-10;
Matches 53; Conservative 39; Mismatches 75; Indels 15; Gaps 7;
QV 7 SIIIPFNAAVTLQACLGISGVIGTYREVEVWLVDGGSTDRDTLDIANSFRPELGSRLVWHS 66

67 GPDDGPDYAMNRGVGVATGEWVLFGLGADDTLYEPTTLAQ--VAAFPLGDHAAASHLVYGDVVM 125
 : ||||| ||| ::| : :
64 QANOGGGAGNAGFAAATGEWVAFLDADD-VMKPHKLQRIVEVFQTSDDVVGVMMHLDIID 122
 : ||||| ||| ||| ||| |||
126 RSTK-----SRHAGP----FDLDRLLPETN---LCH---QSIFYRRELEFDGIGIPYN-LRVRVW 172
 : ||| ||| ::| : :
123 GNDKTIDQASTQGPKLSEDLASVILQTGNACFPPTSGLAYRRREVLEKVPFDPVKRWII 182

173 AD 174
 ||
183 AD 184

RESULT 15
AD2189
hypothetical protein alr3067 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;pare: 14-Dec-2001 #sequence_revision 4-Dec-2001 #text_change 09-Jul-2004
C;accession: AD2189
R;kaneko, T.; Nakamura, Y.; Wolk, C. P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;
Yasuda, M.

;
PRIOR APPLICATION NUMBER: 6

; PRIOR FILING DATE: 2000-11-

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62545
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62545

Query Match      56.2%; Score 788; DB 15; Length 256;
Best Local Similarity 60.9%; Pred. No. 4e-79;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY  1  MTAPVFSIIPTFNAAVTLQACLSIGVGTQYREVEVVLVDGSGTDTLTDIANSPFPELGS 60
Db  1  MAAPFMSIIPTLVAVALPACLDLSIARQTCGDFELVLDGSGTDTLTDIANIFAPNLGE 60

QY  61  RLNVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLVY 120
Db  61  RLIIHRDTDQGVYDAMNRGVGVATGCTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 120

QY  121  GDVVMRSKSHAGFPDLDRLFFETNLCQSIFYYRRELFQDGIQPNLYRYVWADWDFNIR 180
Db  121  GDVIMRSTNFRWGAFLDRLFLKFNICHOAIFYRRGLFGTIGTIPNLYRYVLAOWDFNIR 180

QY  181  CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFKRKLPMYFWVAGWETCRRMLAFKLD 240
Db  181  CFSNPALVTRYMHVVVASYNEFGGLS-NTIVDKFELKRLPM----- 220

QY  241  KENRRLALRTRLIRVK 256
Db  221  --STRLGIRLVILVR 234
```

```
RESULT 2
US-10-282-122A-64808
; Sequence 64808, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64808
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64808

Query Match      56.2%; Score 788; DB 15; Length 275;
Best Local Similarity 60.9%; Pred. No. 4.4e-79;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY  1  MTAPVFSIIPTFNAAVTLQACLSIGVGTQYREVEVVLVDGSGTDTLTDIANSPFPELGS 60
Db  20  MAAPFMSIIPTLVAVALPACLDLSIARQTCGDFELVLDGSGTDTLTDIANIFAPNLGE 79

QY  61  RLNVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAASHLVY 120
Db  80  RLIIHRDTDQGVYDAMNRGVGVATGCTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 139

QY  121  GDVVMRSKSHAGFPDLDRLFFETNLCQSIFYYRRELFQDGIQPNLYRYVWADWDFNIR 180
Db  140  GDVIMRSTNFRWGAFLDRLFLKFNICHOAIFYRRGLFGTIGTIPNLYRYVLAOWDFNIR 199

QY  181  CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFKRKLPMYFWVAGWETCRRMLAFKLD 240
Db  200  CFSNPALVTRYMHVVVASYNEFGGLS-NTIVDKFELKRLPM----- 239

QY  241  KENRRLALRTRLIRVK 256
Db  240  --STRLGIRLVILVR 253
```

```
RESULT 3
US-10-282-122A-78093
; Sequence 78093, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78093
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78093

Query Match 17.0%; Score 238.5; DB 15; Length 247;
Best Local Similarity 32.2%; Pred. No. 8.3e-18;
Matches 65; Conservative 41; Mismatches 75; Indels 21; Gaps 6;

Qy 7 SIITPFNAAVTLQACLGSIIVGQTYREVEVWLVGSGSTDTLTI--ANSFPELGSLV 64
Db 4 SIITAYNSERTISDTLSLEKQTYLDVBYIIVDGASNDTUTVTSQSTRV---TKIIS 60

Qy 65 HSGPDDGPDYDAMNMGVATGEWVLFGLGADDTLYBPTTLAQVAAPLGDHAAASHLVYGDV- 123
Db 61 ES--DQGIYDALNKGIDLATDIIPLGHSDDLAYPGALSIDIVTF-EKQCCDAVYGDQ 117

Qy 124 -----VMRSTKSRHAGPDLDRLKLLPETNLCHQSTFYRRELFDGIGPNLYRVWAD 174
Db 118 YVAKNDLNVKIRCWKS--GFFNHEKQYGMPPHPTFYMKRDLIYRFGGDLKYISAD 174

Qy 175 WDFNIRCFSPALITRYMDVVI 196
Db 175 YDSLTRYIIINYKIKIAYLPKVI 196

RESULT 4
US-10-282-122A-51835
; Sequence 51835, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51835
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51835

Query Match 13.7%; Score 192; DB 15; Length 336;
Best Local Similarity 23.6%; Pred. No. 2e-12;
Matches 65; Conservative 57; Mismatches 113; Indels 40; Gaps 10

Qy 4 PVFSIIPTFNAAVTLQACLGSIIVGQTYREVEVWLVGSGSTDTLTIANSFPELGSLV 63
Db 5 PKVISVMPVNSRYLAELIESILDQTYNDFEIIVDGSGTDESYNIISSYANK-DNRII 63

Qy 64 VHSPPDDGPDYDAMNMGVATGEWVLFGLGADD-----TYEPTTLAQVAAPLGDHAA 115
Db 64 VISREHRLGLVDSINEGINIARGKYIARMDDADDSINNRIEKQFEFLELNKDVLDLGTRE 123

Qy 116 SHLVYGDVWVRST--KSRHAGPDLRL--LPETN--LCHOSIFYRRELFDGIGPNLYR 168
Db 124 A---FGDIEBKQKTYNSAFSKFDSQNIQVFLTSCALPHSPVMPKKDSIVKLRGYRKE 180

Qy 169 YRVWADWDFNIRCFSPALITRYMDVVISEYN-----DMTGFSNRQC----- 210
Db 181 YDTAEDYDLWLRAIRNGYKIVR-MDECLIKYRVHNKSKTAVENFNPKVVEYTMKAKIDVI 239

Qy 211 TDKEFRKRLPMYFVWA--GWETCRRMLAFKKKEN 243
Db 240 NDTNKKDKVDYLWLGASTGGKLVKVVESTTDKFN 274

RESULT 5
US-10-264-213-152
; Sequence 152, Application US/10264213
; Publication No. US20040009490A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Lubbers, Mark William
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them, and methods for using them.
; FILE REFERENCE: 11000.1043C3
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-152

Query Match 13.5%; Score 189.5; DB 15; Length 331;
Best Local Similarity 28.2%; Pred. No. 3.7e-12;
Matches 62; Conservative 38; Mismatches 93; Indels 27; Gaps 9

Qy 4 PVFSIIPTFNAAVTLQACLGSIIVGQTYREVEVWLVGSGSTDTLTIANSFPELGSLV 63
Db 7 PLVSIIVPINVEKYLQRCIDSLFAQTYNIEIIVLVDGSGADRSLSICKK-NAQQDERVR 65

Qy 64 VHSPPDDGPDYDAMNMGVATGEWVLFGLGADDTLYBPTTLAQVAAPL-GDHAASHLVYGD 122

```

Db 66 VFSTKNGGVADTRNFGVSGVQWISFVDPDDYV--DPDVIEYLYGLVKSNAATSGICQHR 124
QY 123 VVMRSTK---SRHAGPFDLRLPETNLCHQSFYRRELFDGIGPYNLYRWAD----W 175
Db 125 NVYKNGKIQTNLVEGPAVLDD-----SHTAV--KRLLYDD-----QIDTSVWAKLYPAW 170
QY 176 DNIRCFNSPALITRYMDVVISYNDMTGFSMRQGTDKEF 215
Db 171 VPKJHFPKGRLL---FEDIAATYKTFLASDSIAVGSEAKY.207

RESULT 6
US-10-282-122A-52169
; Sequence 52169, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-08-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52169
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-52169

Query Match 13.5%; Score 188.5; DB 15; Length 333;
Best Local Similarity 25.8%; Pred. No. 4.8e-12;
Matches 59; Conservative 42; Mismatches 95; Indels 33; Gaps 7;
QY 5 VFSIIPTNAAVTIQAICLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 64
Db 1 MISVMPVYCKEYLEESTESILKQTYRDFEIIVNDGSDKSIDINKYANDNRIVV 60
QY 65 HSGPDDGPDVDMNRGVGATGWLFLGADDTLYEPTTLAQVAAFGLGH-----AASHL 118
Db 61 SRDNNMGVYSLNIGIDRAKGSYVARMDDADD-IALPERFERQIEYLNNKNOVDILACKVE 119

QY 119 VYGDVWRSTKSR-HAGPFDLD-----RLLEFETN--LCHQSIFYRRELFDGIGPYNLYR 170
Db 120 AFGDVSRQKLEREHVYVNDLNNSESTESLENCYIAHPVMMKMSVLKALGGYNLYK 179
QY 171 VWADWDFNIRCFNSPALITRYMDVVIS-----YNDMTGFS 206
Db 180 RTEDYNLWLR-----ATAKGYKIAMLEBKMKIRLHNSDKIHRDAEGFS 223
RESULT 7
US-10-282-122A-66207
; Sequence 66207, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66207
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66207

Query Match 13.4%; Score 188; DB 15; Length 299;
Best Local Similarity 29.8%; Pred. No. 4.7e-12;
Matches 61; Conservative 33; Mismatches 91; Indels 20; Gaps 7;
QY 3 APVFSIIPTNAAVTIQAICLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRL 62
Db 16 APLVSVVAPCFNABKYLEALRSIYEQDYPNFVEIIVDDGSTDNSYAMLEQLQKVHGPQL 75
QY 63 VVHSGPDDGPDVDMNRGVGATGWLFLGADDTLYEPTTLAQVAAFGLGH-----AASH 117
Db 76 --YKQNGVGSAAINFLGRHARGDYVATPDDLDLML-PHLSVRAAYLDQHPVEGCVGAL 132
QY 118 LVYGDVWRSTKSRHAG---PFDLRLLEFETNLCHQSI-FYRRELFDGIGPYNLYR 173


```

Db      133  VIYIDSEGETKRONRIRQDFDYLLGNAYVCGAPVSLYRMEALRAAGFYDPEIKV-Q 191
Qy      174  DMDFNTRCFSN-----PALITRY 191
         | : | | |
Db      192  DPQMTLRIASQGYIHKLPLVTRY 216

```

```

RESULT 8
US/10-182-960-13
; Sequence 13, Application US/10182960
; Publication No. US20030186392A1
; GENERAL INFORMATION:
; APPLICANT: Triemly, Janine, et al.
; TITLE OF INVENTION: Biopolymer Thickner
; FILE REFERENCE: 64376
; CURRENT APPLICATION NUMBER: US/10/182,960
; CURRENT FILING DATE: 2002-08-03
; PRIOR APPLICATION NUMBER: PCT/US01/03404
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/241,098
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/179,888
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Lactococcus lactis
US-10-182-960-13

```

Query Match	13.2%;	Score 184.5;	DB 14;	Length 332;
Best Local Similarity	22.9%;	Pred. No. 1.3e-11;		
Matches 67;	Conservative 63;	Mismatches 111;	Indels 51;	Gaps 13;
Qy	5	VFSIIITFNAAVTLOACLSIVGQTREYEVVLVDGSGTDRITLIDIANSPFELGSLRVV	64	
Db	5	LISIIVPVWSEKYLRAAIHSLNQYINIEVLINDSGTDSQELISSFQKK-DKRIKL	63	
Qy	65	HSGPDDGPGYDAMRGVGTACEWVFLGADDTLYEPTTLAQVAFAFLGDHAASHLYVGDVV	124	
Db	64	YTNKLVGSHARNYGIIDRASGYIMFLDPDOT-YDKSYCLEMIGLKNKFNA-----DW	116	
Qy	125	MRS-----TKSRHAGP-----FDLRLILFTNLCHQSI-----FYRELFQDIG	163	
Db	117	MSNYIICGKNYIPVNNNDLLECGLLSRDKTMSILSDTGFGKFWTRIFRKNVINNV-	175	
Qy	164	PYNLYRYVWADWDENIRCFSPALIT-----RYMDVVISEYNDMTGFS-----MRQG	210	
Db	176	KPNESINYLEDMLEISIVHNARIIVTNKHYPYLORESASKKFSPPKSLNLRGK	235	
Qy	211	TDKEFRKRLP--MYFPWVAGMETCRMLAFKLCKENRRLALRILIRYKAVSK	260	
Db	236	VDEPFYSQIDSVIPYNLVGLMITER-----KSRNSOF-IRRNITKNKMSOVK	281	

RESULT 9
US-10-282-122A-59156
; Sequence 59156, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

, TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
, FILE REFERENCE: ELITRA.034A
, CURRENT APPLICATION NUMBER: US/10/282,122A
, CURRENT FILING DATE: 2003-02-20
, PRIOR APPLICATION NUMBER: 60/191,078
, PRIOR FILING DATE: 2000-03-21
, PRIOR APPLICATION NUMBER: 60/206,848
, PRIOR FILING DATE: 2000-05-23
, PRIOR APPLICATION NUMBER: 60/207,727
, PRIOR FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: 60/230,335
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/230,347
, PRIOR FILING DATE: 2000-09-09
, PRIOR APPLICATION NUMBER: 60/242,578
, PRIOR FILING DATE: 2000-10-23
, PRIOR APPLICATION NUMBER: 60/253,625
, PRIOR FILING DATE: 2000-11-27
, PRIOR APPLICATION NUMBER: 60/257,931
, PRIOR FILING DATE: 2000-12-22
, PRIOR APPLICATION NUMBER: 60/267,636
, PRIOR FILING DATE: 2001-02-09
, PRIOR APPLICATION NUMBER: 60/269,308
, PRIOR FILING DATE: 2001-02-16
, Remaining prior Application data removed - See File Wrapper or PALM.
, NUMBER OF SEQ ID NOS: 78614
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 59156
, LENGTH: 291
, TYPE: PRT
, ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59156

```

Query Match	13.0%;	Score 182;	DB 15;	Length 291;
Best Local Similarity	23.5%;	Pred. No. 2.1e-11;		
Matches 61; Conservative 49; Mismatches 134; Indels 16; Gaps 8;				
Qy	4	PVFSIIPTFNAATLQACLSGVGTQYRVEVVLDGSGTDRDTLDIANSFRPELGSRLV	63	
Db	5	PLVSVPVPTHNAELLSSNAVKSVINQTYKIEIIICNDGSGDNTDDVVNKLQKEYKDIPF	64	
Qy	64	VHSGPDGDPYDA--MNRGVGVATGEVWLFLGADDTLYEPTTLAQVAAPLG----	116	DHAAS
Db	65	VYL-KNDSPMGACFSRNRCTAAASGYLTGLD-DDDYFLPNRIESLKVNCNERKIDL	122	VCS
Qy	117	HLVYGDVVMRSTKSRHAGPPDLORLLFETNLCHQSIFYRRELFDGIGYNNLYRVVADWD	176	
Db	123	NLIFKDGKLLKRGKNSGVITSIDMGVE-NLVGNQLLTRLSYFEDVGGFDTNFPQAWQYD	181	
Qy	177	--FNIRCFSPALITRYMDVWVISEYNDWTCFSMRQGTDKSFRKRLPMYFWVAGHETCRM	234	
Db	182	LWYRIIMKFGPCIKTNEATYVMDVENDRKRISTSKAHIGYRSFIEKHSETLSEB--MKS	239	
Qy	235	LAFKLD--KENRRALALRRL	252	
Db	240	SLFIRDLLINRNKKIPIKLTLL	259	

```

RESULT 10
US-10-282-122A-69114
; Sequence 69114, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

```

```
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69114
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-10-282-122A-69114

Query Match      13.0%; Score 182; DB 15; Length 327;
Best Local Similarity 20.8%; Pred. No. 2.5e-11;
Matches 70; Conservative 46; Mismatches 93; Indels 128; Gaps 13;

QY 1 MTAPVFSIIPTFNAAVTLQACLSIGQTVREVEVLVDGGSTDRDLDIANSFRPELGS 60
Db 1 MKTPIISVIVPFNEASRITRLESLLCQTFHDFEVIINDGSTDINSVDIAMLYCQ-QDN 59
QY 61 RLWHSIGDDGPDYDAMRGVGVATGEWVLFLGADDTLYEPTTLA-----QVAAPFLG 111
Db 60 RPHLYHTNQGLSSARNTGLKYAQGDWIVFDSDDFI-KPQLLAHWHQLACEQHIDVLIG 118
QY 112 D-----HAASHLVYGDVVMRSTKSRHA 133
Db 119 NGERVDVNPQKHQTTIHQRQYQQVISQGEWVHIAVTQHQPFWLQFIRHEILKXH- 177
QY 134 GPFDLRLLFETNLCHQSIFYRREL-----FDGIGPYNLRVYRWADWDFNRCFNSP 185
Db 178 -----HLRFIDGLYHEDILMTTQLALVTQIGFD-----DQPLYYYC-ANP 217
QY 186 ALITR-----YMDVVI-----SEYNDMTGFSNRQGTDXE-----FKR 218
Db 218 NSITRKNPQKBAKQAQSYLIQVILQLLQADKQHQPLLALRQQAALREFGSGFFILFRKR 277
QY 219 -----LPMYF-----WVAGWETCR 232
Db 278 CDTKNQTTIAQQFRYTHLPALNKAQNWRQRFICR 314

RESULT 11
US-10-282-122A-75381
; Sequence 75381, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangeu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
```

```
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75381
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-75381

Query Match      13.0%; Score 182; DB 15; Length 344;
Best Local Similarity 22.4%; Pred. No. 2.7e-11;
Matches 70; Conservative 45; Mismatches 102; Indels 96; Gaps 10;

QY 7 SIIITPFNAAVTLQACLSIGQTVREVEVLVDGGSTDRDLDIANSFRPELGSRLVHS 66
Db 9 SIIIPLYNAGADFACMASLIAQTWSALEIIIVNDGSTDHSIETIAKHYAEPHYVRLHHQ 68
QY 67 GPDDGPDYDAMRGVGVATGEWVLFLGADDTLYEPTTLAQAFAFLGDHAAASHLVYGDVVMR 126
Db 69 A-NAGASVARNLGLQAATGDYVAFVDADLVYPKMYETLMTWALNDDLDVAQCNDWCVR 127
QY 127 STKSRHA-----GPFDL-----DRLLPET 145
Db 128 --KTCHAWQSITPDLRSTGVLSGPDWLRMALASRTHVVMGVYRREALITDNNITFVP 185
QY 146 NLCHQSIFYRRELFEGIGPYNLRVYRWADWDFNRCFNSPALITTRYMDVITSEY---NDM 202
Db 186 GLHHQDILWSTEVN-----FN-----ATRVRYTEQSLYKYFLHDNS 221
QY 203 TGFSNRQGT-----DKEFRKELPMY---FWAGWETCRMLAFLKD 240
Db 222 VSRLOQGNKNLNYQRHYIKITRLEKLNRYARRIPIYEPFROQITWEALRVCHAVRKE 281
QY 241 KENRRRLALRLRI 253
Db 282 PD---ILTRQRM 291

RESULT 12
US-10-282-122A-72785
; Sequence 72785, Application US/10282122A
```

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Cart, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 72785
LENGTH: 343
TYPE: PRT
ORGANISM: Salmonella paratyphi A
US-10-282-122A-72785
Query Match 12.8%; Score 180; DB 15; Length 343;
Best Local Similarity 22.0%; Pred. No. 4.5e-11;
Matches 69; Conservative 46; Mismatches 102; Indels 96; Gaps 10;
QY 7 SIIIPFNAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTDIANSFRPELGSRLVWHS 66
DB 9 SIIIPLYNAGAFNACMSLIIQATWSALSIIVNGSDHSVEIAKHAHYHPRVRLHQ 68
QY 67 GPDDGPDYDAMNNGVGVATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAASHLVYGDVVMR 126
DB 69 A-NAGASVARNLGLQAATGCVYVAFDADLVYPKMYETLMTWALNDLDDVAQCNDWCVR 127
QY 127 STKSRAH-----GPDFL-----DRLLPET 145
DB 128 --KTGHAWOSIPTDRLRSTGVLSGPDWLRMALASRRWTHVVMGVYRRLAIIDNNITFVP 185
QY 146 NLCHOSIFVRRLEFDGIGPNLYRVWADWDNIRCFNSPALITRYMOVVISEY---NDM 202
DB 186 GLHQDILWSTEV-----FN-----ATRVRYTEOSLYKYLHNS 221
QY 203 TGFMSRQGT-----DKFRKRLPMY---FWVAGWETCRMLAFLKD 240
DB 222 VSLRQGNKNLNYQRHYIKITRLLEKLNDRYARRIPIYEPFRQITWEALRVCHAVRKE 281
QY 241 KENRLALRLI 253

Db 282 PD---ILTRQRM 291
RESULT 13
US-09-767-041-36
Sequence 36, Application US/09767041
Patent No. US20020055168A1
GENERAL INFORMATION:
APPLICANT: Smith, Hilda
TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
FILE REFERENCE: 2183-4726
CURRENT APPLICATION NUMBER: US/09/767,041
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: PCT/NL99/00460
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: EP98202465.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: EP98202467.1
PRIOR FILING DATE: 1998-07-22
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 36
LENGTH: 278
TYPE: PRT
ORGANISM: Streptococcus suis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: CPS1K
US-09-767-041-36
Query Match 12.8%; Score 179; DB 9; Length 278;
Best Local Similarity 23.6%; Pred. No. 4.3e-11;
Matches 61; Conservative 47; Mismatches 93; Indels 59; Gaps 10;
QY 2 TAPVFIIIPFNAVTLQACLSIVGQTYREVEVVLVDGSGTDTLTDIANSFRPELGSR 61
DB 3 TISKIISIIPIYNVEKYSKCIDSIIVNQTYSKHIELLVNDGSDTNDSEIEICLAYAKK-DSR 61
QY 62 LVVHSGPDGPDYDAMNNGVGVATGEWVFLGADDTLY-----EPTTLAQVAAP 109
DB 62 IRYFKKENGSLSDARNYGISRAKGYLAPIDSDDDFIHSEFIORLHEATERENALVAVAGY 121
QY 110 LGDHAASHLVYGDVVMRSTKSRHAGFPDLDRLLFTNLC-----HQSI-----FYRR 156
DB 122 DRVDASGHFLTAEP LPTN-----QAVLSGRNVCKKLLLEADGHRFVVACNKLYKK 170
QY 157 ELFDGIGPNLYR---RVWADWDNIRCFNSPALITRYMDVVISEVNDMTGSMRQG 210
DB 171 ELFE-----DFRFEKGKIHEDYEFTYRLLYELEKVAIVKBCLYYYVDRENSITTSM--- 222
QY 211 TDKEP-----RKRLPMY 222
DB 223 TDHREHCLLEFQNERMDFY 241
RESULT 14
US-10-282-122A-67095
Sequence 67095, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Cart, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67095
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67095

Query Match      12.8%; Score 179; DB 15; Length 477;
Best Local Similarity 21.6%; Pred. No. 9.2e-11;
Matches 59; Conservative 54; Mismatches 98; Indels 62; Gaps 7;

QY 1 MTAPVFSIIPTFNAAVTLOACLGSIYGQTYREVEVVLVDGSGTDRDLDIANSFPRLGSGS 60
Db 211 LNAPLITILVTTFNSQKSIKNTLSNFQSPYNIEIIVIDHSQDNTWSILQAYTKQYKN 270

QY 61 RLIVHSGPDGPDYDAMNRGVGVATGEWVLFGLGADDTLYEPTTTLAQVAAFLGDHAAASHLVY 120
Db 271 IKIISLKENGVYVAKNIGLKYASGEFFTCQSDDDMAHPQKLALQVAPL-----LQH 322

QY 121 GGVNMRSTKSRHAGP-----FDLRLLPETNLCHQSIFYRRELFDGIGPYNLRVYV 171
Db 323 KELIVTFKSWVRLEDPIGNPYARTIYPLMLNPSS-----ALPRKEVCE-----KTAL 370

QY 172 WADWDFNIRCFSPNALITRYMDVVISYNDMTGFSMROGTDKEFRKRLPMYFWVAGWETC 231
Db 371 W-DW-----VRIGADSEFNARUKLIFGHKGYYTV 398

QY 232 RRLAFLKDKENRRRLALRTLIRVKAVERSA 264
Db 399 NKPLTFGAHRENSLMTAQSTGY-VNGVSLPREA 430

RESULT 15
US-09-767-041-22
; Sequence 22, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22

; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CFS2K
US-09-767-041-22

Query Match      12.7%; Score 178; DB 9; Length 332;
Best Local Similarity 21.9%; Pred. No. 7.1e-11;
Matches 68; Conservative 53; Mismatches 100; Indels 90; Gaps 12;

QY 7 SIIITFNAAVTLOACLGSIYGQTYREVEVVLVDGSGTDRDLDIANSFPRLGSLVHVS 66
Db 5 SIIIVINVEQYLSKCINSIVNQTYKHIEILLVNDGSTDNSEEICLAYAKK-DSRIRYFK 63

QY 67 GPDDGPDYDAMNRGVGVATGEWVLFGLGADDTLY-----EPTTLAQVAAFLGDHA 114
Db 64 KENGGLSDARNYGISRAKGDYLAFLIDSDDFIHSBFQRLHEAIRENALVAVAGYDRVDA 123

QY 115 ASHLVYGDVVMRSTKSRHAGPDLRLLFETNLC-----HOSI-----FYRRELFDG 161
Db 124 SCHFLTAEPPLPTN-----QAVLSGRNVCKLLEADGHRFVAVWANKLYKKELED- 171

QY 162 IGPYNLRY---RVWADWDFNIRC---FSNPALITRYMDVVISYNDMTGFSMROGTDKEF 215
Db 172 -----FRFEKGIHEDEVFTYRLVLELEKVAIVKESCLVYVVDRENSIITSSM---TDHRF 223

QY 216 -----RKRLPMY-----FWVAGHETCERML---AF 237
Db 224 HCLLEFQNERMDFYESRDKELLECYSRFLAFAVLFLGKYNHMLSKOOKKLQTLFRIVY 283

QY 238 LKDKENRRRLAL 248
Db 284 KQLKQNKRLAL 294

Search completed: February 27, 2005, 21:23:23
Job time : 136 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 21:02:03 ; Search time 43 Seconds
(without alignments)
461.783 Million cell updates/sec

Title: us-10-805-311-24

Perfect score: 1401

Sequence: 1 MTPVFSIIITFTNAATLQ.....ALRTRLIRKAVKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	56.2	274	4	US-08-311-731A-16
2	189.5	13.5	331	4	US-09-634-238-247
3	188	13.4	303	4	US-09-252-991A-29155
4	178	12.7	340	4	US-09-543-681A-8334
5	174	12.4	330	4	US-09-134-000C-6392
6	172	12.3	333	4	US-09-107-532A-5123
7	164.5	11.7	842	4	US-09-134-000C-5749
8	163.5	11.7	534	4	US-09-134-000C-5087
9	163	11.6	721	4	US-09-107-532A-6889
10	161	11.5	324	1	US-08-597-236-10
11	161	11.5	324	1	US-08-746-682A-10
12	158	11.3	348	1	US-08-312-387B-3
13	158	11.3	348	1	US-08-312-387B-11
14	158	11.3	348	1	US-08-683-426-3
15	158	11.3	348	1	US-08-683-426-11
16	158	11.3	348	1	US-08-683-458-3
17	158	11.3	348	1	US-08-683-458-11
18	158	11.3	348	2	US-08-878-360-3
19	158	11.3	348	2	US-08-878-360-11
20	158	11.3	348	3	US-08-478-140B-3
21	158	11.3	348	3	US-08-478-140B-8
22	158	11.3	348	3	US-09-333-412-3
23	158	11.3	348	3	US-09-333-412-11
24	158	11.3	348	3	US-09-338-943-3
25	158	11.3	348	3	US-09-338-943-8
26	158	11.3	348	4	US-10-007-267A-3
27	158	11.3	348	4	US-10-007-267A-11

28	156.5	11.2	270	4	US-09-495-406-25
29	156.5	11.2	270	4	US-09-816-028A-39
30	156.5	11.2	270	4	US-10-303-162-39
31	156.5	11.2	270	4	US-10-303-134-39
32	155.5	11.1	376	4	US-09-902-540-13915
33	154.5	11.0	301	4	US-09-252-991A-33096
34	151.5	10.8	328	4	US-09-583-110-3554
35	151.5	10.8	331	4	US-09-107-433-4710
36	151	10.8	327	4	US-09-107-532A-6181
37	150	10.7	187	4	US-09-973-457-4
38	149.5	10.7	703	4	US-09-902-540-11194
39	149.5	10.7	972	4	US-09-489-200E-9
40	147.5	10.5	965	4	US-09-437-277-3
41	147	10.5	750	4	US-09-489-039A-8940
42	146.5	10.5	301	4	US-09-495-406-15
43	146.5	10.5	301	4	US-09-816-028A-27
44	146.5	10.5	301	4	US-10-303-162-27
45	146.5	10.5	301	4	US-10-303-134-27

ALIGNMENTS

RESULT 1
US-08-311-731A-16
; Sequence 16, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: /08/311,731A
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
US-08-311-731A-16

Query Match 56.2%; Score 788; DB 4; Length 274;

Best Local Similarity 60.9%; Pred. No. 1.5e+83;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

Qy 1 MTPVFSIIITFTNAATLQCLGSIVGQTYREVVLDGGSTDRTLDIANSFRPELGS 60

Db 19 MAAPNFSLIIPLNVAAVLPACLDSIAQTCGDFELVLVDGGSTDTETLIDIANIFAPNLGE 78
Qy 61 RLIVHSGDDPGYDAMNRGVGATGEWVFLFGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
Db 79 RLIIHRDTDQGVYDAMNRGVDLATGTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 138
Qy 121 GDVVRSTKSRHAGPFDRLDLFFETNLCHQSIFVYRRELFDGIGPNLYRYVWADWFNIR 180
Db 139 GDVNRSTNFRWGGAFDRLDLFFKNICHQAIFVYRRLFGTIGPNLYRYVLAOWDFNIR 198
Qy 181 CFSNPALITRYMDVVISYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCKRMLAFKLD 240
Db 199 CFSNPALVTRYMHVVVASYNEFGLS-NTIYDKEFLKRLPM----- 238
Qy 241 KENRLAURTLRIYK 256
Db 239 --STRLGIRLVILVR 252

RESULT 2
US-09-634-238-247
; Sequence 247, Application US/09634238
; Patent No. 6544772

GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna C.
; APPLICANT: Hølland, Ross
; APPLICANT: O'toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Cbolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.

FILE REFERENCE: 11000.1043UI
; CURRENT APPLICATION NUMBER: US/09/634,238
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 247
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-247

Query Match 13.5%; Score 189.5; DB 4; Length 331;
Best Local Similarity 28.2%; Pred. No. 2e-13; 93; Indels 27; Gaps 9;
Matches 62; Conservative 38; Mismatches 93;
Qy 4 PVFSIIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 63
Db 7 PLVSIIVPIYVNEKYLQRCIDSLPAQTYVNIELVLVDGSDADSLICK-NAQQDERVR 65
Qy 64 VHSFGDDPGYDAMNRGVGATGEWVFLFGADDTLYEPTTLAQVAAFL-GDHAASHLVYGD 122
Db 66 VFSKTNGVADTRNFGVSVARGWISFVDPDDYV-DPDYIYLYGILVKSNAWTMSICOHR 124
Qy 123 VMVRSTK---SRHAGPFDRLDLFFETNLCHQSIFVYRRELFDGIGPNLYRYVWAD---W 175
Db 125 NVYKNGKIQTNLYEGPAVL-----SHYAV--KRLLYDD----QDTSVWAKLYPAW 170
Qy 176 DFNIRCFSPALITRYMDVVISYNDMTGFSMRQGTDF 215
Db 171 VFVKTHFPKGRLL--FEDIAATYKTFILASDSIAVGSEAKY 207

RESULT 3
US-09-252-991A-29155
; Sequence 29155, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29155
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29155

Query Match 13.4%; Score 188; DB 4; Length 303;
Best Local Similarity 29.8%; Pred. No. 2.6e-13;
Matches 61; Conservative 33; Mismatches 91; Indels 20; Gaps 7;
Qy 3 APVFSIIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRL 62
Db 20 APLSVVAPCFNAEKYLEEARLSIYEQDYPNFEVLIIVDDGSTDNSYAMLEQLQKVHGFQL 79
Qy 63 VHSFGDDPGYDAMNRGVGATGEWVFLFGADDTLYEPTTLAQVAAFLGDH-----AASH 117
Db 80 --YRQONQGVSAALNFGLOHARGDYATPDLDLIML-PHSLSVRAAYLDQHPVGCVGAL 136
Qy 118 LVYGDVVMRSTKSRHAG---PFDLRLFFETNLCHQSI-FYRRELFDGIGPNLYRYVWA 173
Db 137 VIYIDSEGTETKQNGNRIRQLDFYLLGNAYVCGAPSVLYRMEALRAAGFYDPEIKV-Q 195
Qy 174 DWDFNIRCFSN-----PALITRY 191
Db 196 DFQMTLIASQGYQIHKLPVLVTRY 220

RESULT 4
US-09-543-681A-8334
; Sequence 8334, Application US/09543681A
; Patent No. 6605709
GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8334
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8334

Query Match 12.7%; Score 178; DB 4; Length 340;
Best Local Similarity 20.7%; Pred. No. 4.6e-12;
Matches 69; Conservative 46; Mismatches 91; Indels 128; Gaps 13;
Qy 4 PVFSIIIPTFNAAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLV 63
Db 17 PIISVIVMFENEAASRITLLESLLCQTFHDIFEVIINDGSTDNSVDIAMLYCQO-DNRFH 75
Qy 64 VHSFGDDPGYDAMNRGVGATGEWVFLFGADDTLYEPTTLA-----QVAAFLGD-- 112
Db 76 LYHOTNOGLSSARNAGLUKYGQDWIVFDSDDFI-KPOLLAHWHQLASEQHIDVLINGGE 134
Qy 113 -----HAASHLVYGDVVMRSTKSRHAGPF 136

Db 135 RYDVNPKQHTIHRQPOYQOVISQEWVIHANTQHPHVMQLFIRHBIKKH----- 190
Qy 137 DLDRLLFETNLCHQSFYRREL-----FDGIGPNLYRVWADWDFNRCFSNPALI 188
Db 191 ---HLRFIDGLYHEDILMTTQLALVTQRIGD-----DQPLYVC-ANFNSI 233
Qy 189 TR-----YMDVVI-----SEYNDMTGFSMRQGTDKB-----FRKR--- 218
Db 234 TRKNPQKEAKAQSQYLQIVITQLQLADKQHPQLANALROALRELGSFPLFRKRCDT 293
Qy 219 -----LPMVF-----WVAGMETCR 232
Db 294 KQALIAQOQSTYHLFPALNKGAQNRORWFICR 327

RESULT 5

US-09-134-000C-6392
; Sequence 6392, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6392
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6392

Query Match 12.4%; Score 174; DB 4; Length 330;

Best Local Similarity 23.6%; Pred. No. 1.3e-11;
Matches 68; Conservative 49; Mismatches 99; Indels 72; Gaps 11;

Qy 4 PWFSTIIITFNNAVTLQACLSIVGQTYREVEVLVDGSGDRTLDIANSFRPELGSRLV 63
Db 8 KRISIIIVPVNVKYLEKCVRSILAQTTDFELILVDGSPDSSGAMCDQF-AEQDQKVK 66
Qy 64 VHSQPDGPDYDMNRGCVATGEWVFLGAD-----DTLYEPTTLAQVAAPLGDHAAHLV 119
Db 67 VIHKGSLDARNAGIATGEYLGFDSDDYIADMYE-----LL 108
Qy 120 YGDVVMRS-----TKSRHAGPDLDR---LLFETNLCH---QSIFVR 155
Db 109 YTNIVKEDADLSICGIYDVYEGKEPIVKSLIQGTFSEALLILQGNIIISVHVAVKLYK 168
Qy 156 RELFPGIGPNLYRVWADWDFNRCFSNPALITRYMDVISE-----YNDMTGFSMRQ 209
Db 169 RKLFP-----ADLRY---PKGTHDESPFIVDLLSECQKVSIDSTQKYYHYHRMGSINTET 220
Qy 210 GTDKFRKRLPMYFVAGWTCRRML-----AFKDKENKRLALRRLI 253
Db 221 FSDKQFE-----FIEAWKRLKKGAVIEAAHQVCFANFLV 261

RESULT 6

US-09-107-532A-5123
; Sequence 5123, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street

CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5123:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...333
; SEQUENCE DESCRIPTION: SEQ ID NO: 5123:
US-09-107-532A-5123

Query Match 12.3%; Score 172; DB 4; Length 333;

Best Local Similarity 27.5%; Pred. No. 2.3e-11;
Matches 52; Conservative 33; Mismatches 56; Indels 48; Gaps 6;

Qy 7 SIIITPTNAAVTLQACLSIVGQTYREVEVLVDGSGDRTLDIANSFRPELGSRLV 66
Db 14 SIIIVPVNVKYLEKCVDSILGQTFDFELLIDGSDTNSGICDEL-AKTDNRKIVH 72
Qy 67 GPDDGPDYDMNRGCVATGEWVFLGADDTLYEPTTLAQVAAPLGDHAAHLVGDVVR 126
Db 73 KENGGLDARNIGIEVAKGDFIGFIDSDDYIDE-----DMYAFI-----YNNILKY 118
Qy 127 STKSRHAGPFD-----LDRLLFETNLCHQSFYRRELFD 160
Db 119 DAELSMGCIYDVHKNKEIKLTPPSQLVTKEALIELVDGKLIVVANAV--SKLYKELFE 176
Qy 161 GIGPNLYR 169
Db 177 -----NVRY 180

RESULT 7

US-09-134-000C-5749
; Sequence 5749, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15

QY	173	ADWDFNIRCFNPALITRYMDVISEYNDMTGFSNRQGTDFKFRKRLP--MYFW	224
Db	367	QDYDLVLR-----FTEKTKERIKHPIKVLVYV	394
RESULT 9			
US-09-107-532A-6889			
; Sequence 6889, Application US/09107532A			
; Patent No. 6583275			
; GENERAL INFORMATION:			
; APPLICANT: Lynn A Doucette-Stamm and David Bush			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO			
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS			
; NUMBER OF SEQUENCES: 7310			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION			
; STREET: 100 Beaver Street			
; CITY: Waltham			
; STATE: Massachusetts			
; COUNTRY: USA			
; ZIP: 02354			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: CD-ROM ISO9660			
; COMPUTER: PC			
; OPERATING SYSTEM: <Unknown>			
; SOFTWARE: ASCII			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/107,532A			
; FILING DATE: 30-Jun-1998			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 60/085,598			
; FILING DATE: 14 May 1998			
; APPLICATION NUMBER: 60/051571			
; FILING DATE: July 2, 1997			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Ariniello, Pamela Deneke			
; REGISTRATION NUMBER: 40,489			
; REFERENCE/DOCKET NUMBER: GTC-012			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (781)893-5007			
; TELEFAX: (781)893-8277			
; INFORMATION FOR SEQ ID NO: 6889:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 721 amino acids			
; TYPE: amino acid			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
; HYPOTHETICAL: YES			
; ORIGINAL SOURCE:			
; ORGANISM: Enterococcus faecium			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (B) LOCATION 1...721			
; SEQUENCE DESCRIPTION: SEQ ID NO: 6889:			
US-09-107-532A-6889			
Query Match 11.6%; Score 163; DB 4; Length 721;			
Best Local Similarity 25.0%; Pred. No. 8.4e-10;			
Matches 55; Conservative 37; Mismatches 118; Indels 10; Gaps 6;			
QY	4	PVFSIIPTFNAAVT-LQACLSIVGQTYREVEVLVDGGSTDRDLTDIANSFRPELGSRL	62
Db	187	PLISIMPVYNVEIKWLEKCIDSLVDQTYDHWELCIDSDASTDPAIRKCLSYQAKDRI	246
QY	63	VVHSGDDPGPYD-AMNRGVGATGEWVFLGADDTLYPEPTTLAQVAFLGDHAAHSLVYG	121
Db	247	KVVFQENGHISLATNSALEMAEGEFTALDNDDEL-PPFALYEVAKVLNVHPELDLYS	305
QY	122	DVVMSTKSRHAGP-----FDLRLPETNLCHOSIFYRRELFGICGYPNLYRVVADWD	176
Db	306	DEDKIDADGNRPDPHFKADWSPDTLMGNVYISHLGV-YRTSIVKELGFRKGYEGSQDY	364

QY 177 FNRCSN-PALITRYMDVVISEYNDMTGFSMRQGTDF 215
DB 365 LVLRVTEQIPAEHIVHIDRVLYHWRITPGSTASNGEAKSY 404

RESULT 10

US-08-597-236-10
; Sequence 10, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-597-236-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 4.2e-10;
Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;

QY 5 VFSIIPTFNAAVTLQACLSIGVQTYREVEVLVDGGSTDTLTDIANSFRPELGSRLV 64
DB 6 LISIVIPVNVKYLEKCLQSQVQNTYNNFEVILVNDGSTDSLSICEKFVNO-DKRFV 64
QY 65 HSGPDDGPDYDANRGVGVATGEWVLFLGADDTLYEPTTLAQVAFLGHAASHLYVG 121
DB 65 FSKENGGMSSARNFGIKKAGSFITFVSDDDYI-----VKDYL-----SHLVAG 108

RESULT 11

US-08-746-682A-10
; Sequence 10, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York

COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,682A
FILING DATE: 14-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/597,236
FILING DATE: 20-JUN-1995
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-746-682A-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 4.2e-10;
Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;

QY 5 VFSIIPTFNAAVTLQACLSIGVQTYREVEVLVDGGSTDTLTDIANSFRPELGSRLV 64
DB 6 LISIVIPVNVKYLEKCLQSQVQNTYNNFEVILVNDGSTDSLSICEKFVNO-DKRFV 64
QY 65 HSGPDDGPDYDANRGVGVATGEWVLFLGADDTLYEPTTLAQVAFLGHAASHLYVG 121
DB 65 FSKENGGMSSARNFGIKKAGSFITFVSDDDYI-----VKDYL-----SHLVAG 108

RESULT 12

US-08-312-387B-3
; Sequence 3, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-3

Query Match 11.3%; Score 158; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 1e-09;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGDRTLDIANSRPELGSRLV 63
Db 3 PLVSVLICAYNVEKYFAQSLAAVNVQTNWRNDILIVDDGSDGTGLAIADKFQKDRSRIKI 62
QY 64 VHSQPDGPGPYDAMNRGV-----GVATGEWVFLGADD-----TLYEPTTLAQ 105
Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GDHAASHLVYGVVWRSTKSRHAGFPDLDRLLEFNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEKDGKRLARHHKHGKIWKPKTRHEDIAAF-----PFGNPIHNNTMIMERS 178
QY 158 LFDGIGPYNLYRVWADW 175
Db 179 VIDG-----GLRYDTERDW 192

RESULT 14
US-08-683-426-3
; Sequence 3, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-426-3

Query Match 11.3%; Score 158; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 1e-09;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGDRTLDIANSRPELGSRLV 63
Db 3 PLVSVLICAYNVEKYFAQSLAAVNVQTNWRNDILIVDDGSDGTGLAIADKFQKDRSRIKI 62
QY 64 VHSQPDGPGPYDAMNRGV-----GVATGEWVFLGADD-----TLYEPTTLAQ 105
Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GDHAASHLVYGVVWRSTKSRHAGFPDLDRLLEFNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEKDGKRLARHHKHGKIWKPKTRHEDIAAF-----PFGNPIHNNTMIMERS 178
QY 158 LFDGIGPYNLYRVWADW 175
Db 179 VIDG-----GLRYDTERDW 192

RESULT 13
US-08-312-387B-11
; Sequence 11, Application US/08312387B
; Patent No. 554553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-11

Query Match 11.3%; Score 158; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 1e-09;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;
```

Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGYYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GHAASHLYVDVVMRSTKSRHAGPFDLRLLFETNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEEDGKRNLRHKGKWKPKTRHEDIAAF----PFGNPIHNNTWIMRRS 178
QY 158 LFDGIGPNLYRYRWADW 175
Db 179 VIDG----GLRYDTERDW 192

RESULT 15
US-08-683-426-11
; Sequence 11, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-11

Query Match 11.3%; Score 158; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 1e-09;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;
QY 4 PVFSIIPTFNAVTLQACLSIGVQTYREVEVLVDGGSTDRITLDIANSFPELGSRIV 63
Db 3 PLVSVLICAYNVEKYPAQSIAAVNQTWRNLILIVDDGSTDGTIAIAKDFKDRSRIKI 62
QY 64 VHSQPDGPDYDANRGV-----GVATGEWVFLGADD-----TLYEPTTLAQ 105
Db 63 LAQAQNSGLIPSLNIGLDELAKSGGGGYYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
QY 106 VAAFL-----GHAASHLYVDVVMRSTKSRHAGPFDLRLLFETNLCHQSIFYRRE 157
Db 123 MGAWLEVLSEEDGKRNLRHKGKWKPKTRHEDIAAF----PFGNPIHNNTWIMRRS 178
QY 158 LFDGIGPNLYRYRWADW 175

Db 179 VIDG----GLRYDTERDW 192
Search completed: February 27, 2005, 21:12:23
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2005, 20:48:48 ; Search time 165 Seconds
(without alignments)
623.505 Million cell updates/sec

Title: US-10-805-311-24

Perfect score: 1401
Sequence: 1 MTAPVFSIIPTFNAATLQ.....ALRTRLIRKAVKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1401	100.0	266	2	Aaw21775 Protein e
2	1401	100.0	266	2	Aaw21774 Protein e
3	788	56.2	256	6	Abu34621 Protein e
4	788	56.2	274	7	Abd74267 Mycobacte
5	788	56.2	275	2	Aaw21783 Protein e
6	788	56.2	275	6	Abu36884 Protein e
7	298	21.3	297	2	Aaw88309 E. coli c
8	238.5	17.0	247	6	Abu50169 Protein e
9	205.5	14.7	298	4	Aab96313 Putative
10	202.5	14.5	262	2	Aaw21779 Protein e
11	201.5	14.4	321	5	Abp26805 Streptoco
12	192	13.7	336	6	Abu23911 Protein e
13	190	13.6	248	6	Aaw88314 Sugar tra
14	190	13.6	248	7	Adc01227 Enterohae
15	189.5	13.5	331	8	Adi67157 Lactobaci
16	188.5	13.5	333	6	Abu24245 Protein e
17	188	13.4	299	6	Abu38283 Protein e
18	188	13.4	303	7	Abu80409 Pseudomon
19	186.5	13.3	316	3	Aay43774 Amino aci
20	186.5	13.3	316	3	Aay54072 Enzyme BP
21	186	13.2	277	2	Aaw21767 Protein e
22	184.5	13.2	332	4	Aab47426 EpsN. 10/
23	182	13.0	291	6	Abu31232 Protein e
24	182	13.0	327	6	Abu41190 Protein e
25	182	13.0	344	6	Abu47457 Protein e

26	180.5	12.9	327	5	ABP26804	Abp26804 Streptoco
27	180	12.8	343	6	ABU44861	Abu44861 Protein e
28	179	12.8	278	3	AAV68976	AAV68976 CpsK pro
29	179	12.8	477	6	ABU39171	Abu39171 Protein e
30	178	12.7	340	7	ADF08049	Adf08049 Bacterial
31	177	12.6	421	6	ABU20309	Abu20309 Protein e
32	176	12.6	360	6	ABU20878	Abu20878 Protein e
33	175	12.5	334	3	AAV68963	AAV68963 Cps2K pro
34	174.5	12.5	333	5	ABB53519	Abb53519 Lactococc
35	174.5	12.5	623	5	ABB48553	Abb48553 Listeria
36	174.5	12.5	623	6	ABU32458	Abu32458 Protein e
37	174	12.4	330	7	ADH88507	Adh88507 Enterococ
38	173.5	12.4	331	6	ABU28458	Abu28458 Protein e
39	172.5	12.3	344	6	ABU28820	Abu28820 Protein e
40	172	12.3	333	7	ADC95496	Adc95496 E. faeciu
41	172	12.3	358	4	AAU34218	Aau34218 Staphyloc
42	172	12.3	573	4	AAU36881	Aau36881 Staphyloc
43	172	12.3	573	6	ABU15892	Abu15892 Protein e
44	172	12.3	574	6	ABW71992	Abw71992 Staphyloc
45	170.5	12.2	301	6	ABU02243	Abu02243 S. pneumo

ALIGNMENTS

RESULT 1
AAW21775

ID AAW21775 standard; protein; 266 AA.

AC AAW21775;

XX

DT 17-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX

DE Protein encoded by ORF F of the GS region in M. paratuberculosis.

XX GS; pathogenecity island; pathogenic protein; mycobacterial disease;

KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;

KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.

XX Mycobacterium avium subsp. paratuberculosis.

XX WO9723624-A2.

XX 03-JUL-1997.

XX 23-DEC-1996; 96WO-GB003221.

XX 21-DEC-1995; 95GB-00026178.

XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N; Ford J;

XX WPI; 1997-351061/32.

XX N-PSDB; AAT74470.

XX New isolated pathogenicity island from mycobacteria - used to develop products for detection, diagnosis, prevention and treatment of mycobacteria infections.

XX Claim 1; Page 51; 62pp; English.

XX The present sequence represents a protein encoded by open reading frame (ORF) F of a novel polynucleotide sequence designated "GS". GS is a pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb with multiple ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also the transmissible element, encode proteins which may be linked to pathogenecity, such as providing receptors for cellular recognition. GS was discovered and characterised using differential DNA analysis technology. It is found within Mycobacterium paratuberculosis

CC and it has also been identified in Mycobacterium avium subspecies
 CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
 CC the intestine and Crohn's disease in humans. The protein products of the
 CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
 CC treating or preventing mycobacterial disease. In particular they can be
 CC used as vaccines for inflammatory diseases such as Crohn's disease or
 CC sarcoidosis in humans or Johne's disease in animals. (Updated on 25-MAR-
 CC 2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX Sequence 266 AA;
 SQ Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e-143; Indels 0; Gaps 0;
 Matches 266; Conservative 0; Mismatches 0;
 QY 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGGSTDTLTDIANSFRPELGS 60
 DB 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGGSTDTLTDIANSFRPELGS 60
 QY 61 RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 DB 61 RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDENIR 180
 DB 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDENIR 180
 QY 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 DB 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 QY 241 KENRRALRTRLIIRVKAVERKERSAEP 266
 DB 241 KENRRALRTRLIIRVKAVERKERSAEP 266
 RESULT 2
 AAW21774 ID AAW21774 standard; protein; 266 AA.
 XX AC AAW21774;
 XX 17-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 11-MAR-1998 (first entry)
 XX Protein encoded by ORF F of GS region in M. avium subspecies silvaticum.
 XX GS; pathogenecity island; pathogenic protein; mycobacterial disease;
 KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
 KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.
 XX Mycobacterium avium; subspecies silvaticum.
 XX WO9723624-A2.
 PN 03-JUL-1997.
 XX 23-DEC-1996; 96WO-GB003221.
 XX 21-DEC-1995; 95GB-00026178.
 XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
 PA Hermonataylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
 PI Ford J;
 XX WPI; 1997-351061/32.
 DR N-PSDB; AAT74469.
 XX New isolated pathogenecity island from mycobacteria - used to develop
 PT products for detection, diagnosis, prevention and treatment of

PT mycobacteria infections.
 XX Claim 1; Page 50; 62pp; English.
 XX The present sequence represents a protein encoded by open reading frame
 CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
 CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
 CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
 CC ORFs, and also the transmissible element, encode proteins which may be
 CC linked to pathogenecity, such as providing receptors for cellular
 CC recognition. GS was discovered and characterised using differential DNA
 CC analysis technology. It is found within Mycobacterium paratuberculosis
 CC and it has also been identified in Mycobacterium avium subspecies
 CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
 CC the intestine and Crohn's disease in humans. The protein products of the
 CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
 CC treating or preventing mycobacterial disease. In particular they can be
 CC used as vaccines for inflammatory diseases such as Crohn's disease or
 CC sarcoidosis in humans or Johne's disease in animals. (Updated on 25-MAR-
 CC 2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX Sequence 266 AA;
 SQ Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e-143; Indels 0; Gaps 0;
 Matches 266; Conservative 0; Mismatches 0;
 QY 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGGSTDTLTDIANSFRPELGS 60
 DB 1 MTAPVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGGSTDTLTDIANSFRPELGS 60
 QY 61 RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 DB 61 RLIVHSGPDDGPDYDAMNRGVGVATGEWVFLGADDTLYEPTTLAQVAAFLGDHAAASHLVY 120
 QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDENIR 180
 DB 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHOSIFYRRELFDGIGPNLYRVWADWDENIR 180
 QY 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 DB 181 CFSNPALITRYMDVVISEYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRRMLAFLKD 240
 QY 241 KENRRALRTRLIIRVKAVERKERSAEP 266
 DB 241 KENRRALRTRLIIRVKAVERKERSAEP 266
 RESULT 3
 ABU34621 ID ABU34621 standard; protein; 256 AA.
 XX AC ABU34621;
 XX 19-JUN-2003 (first entry)
 DT Protein encoded by Prokaryotic essential gene #20148.
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Mycobacterium bovis.
 XX WO20027183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA38491.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 62545; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 256 AA;
Query Match 56.2%; Score 788; DB 6; Length 256;
Best Local Similarity 60.9%; Pred. No. 7e-77;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
QY 1 MTAPEVSIIPFTNAAVTQACLSIGVQTVREVEVVLVDGSGDRTLDIANSFRPELGS 60
DB 1 MAAPMFSSIIPTLNVAAPLACLDISARQTCGDFELVLDGSGDRTLDIANSFRPELGS 60
QY 61 RLVVHSGDDPDPYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 120
DB 61 RLIIHRDQDQVYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 120
QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHQISFYRRELFDGIGPNLYRYRWADWDFNIR 180
DB 121 GDVIMRSTNFRWGGAFDLRLFFETNLCHQISFYRRELFDGIGPNLYRYRWADWDFNIR 180
QY 181 CFSNPALITRYMDVVI SEYNDMTGFSMRQGTDFKRLKPLFYFWVAGWETCRRLAFLKD 240
DB 181 CFSNPALVTRYMYHVVVASYNEFGCLG-NTIVDKFKLPLW----- 220
QY 241 KENRRLAIRTLIRVK 256
DB 221 --STRLGRLVILVR 234

RESULT 4
ADB74267
ID ADB74267 standard; protein; 274 AA.
XX
XX ADB74267;
XX
XX 04-DEC-2003 (first entry)
XX
XX Mycobacterium tuberculosis non-naturally occurring peptide #15.
XX
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;
XX hypersensitivity reaction; tuberculostatic.
XX
XX Mycobacterium tuberculosis.
XX
XX US6593266-B1.
XX
XX 24-JUN-2003.
XX
XX 16-SEP-1994; 94US-00311731.
XX
XX 19-AUG-1993; 93US-00109181.
XX
XX 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Mao J;
XX
XX WPI; 2003-656441/62.
XX
XX N-PSDB; ADB74252.
XX
XX New Mycobacterium tuberculosis anion pump peptide useful for as
XX tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX Disclosure; SEQ ID NO 16; 26pp; English.
XX
XX The invention relates to a non-naturally occurring peptide of
XX Mycobacterium tuberculosis comprising an amino acid sequence
XX corresponding to an anion pump protein. The invention also relates to a
XX non-naturally occurring nucleic acid corresponding to a DNA sequence of
XX Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
XX useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
XX leprae or for screening for new tuberculosis drugs. Purified proteins
XX derived from the sequences of the invention may elicit a specific immune
XX response. The peptide may also be used to detect hypersensitivity
XX reactions of individuals exposed to Mycobacterium tuberculosis or
XX Mycobacterium leprae. The proteins and peptides may be affixed to solid
XX supports to detect antibodies typical of hypersensitivity reactions, from
XX a patient's sera. This sequence represents Mycobacterium tuberculosis non
XX -naturally occurring peptide of the invention. Note: The sequence data
XX for this patent did not form part of the printed specification but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 274 AA;

Query Match 56.2%; Score 788; DB 7; Length 274;
Best Local Similarity 60.9%; Pred. No. 7.7e-77;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
QY 1 MTAPEVSIIPFTNAAVTQACLSIGVQTVREVEVVLVDGSGDRTLDIANSFRPELGS 60
DB 19 MAAPMFSSIIPTLNVAAPLACLDISARQTCGDFELVLDGSGDRTLDIANSFRPELGS 78
QY 61 RLVVHSGDDPDPYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 120
DB 79 RLIIHRDQDQVYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 138
QY 121 GDVVMRSTKSRHAGPFDLRLFFETNLCHQISFYRRELFDGIGPNLYRYRWADWDFNIR 180
DB 139 GDVIMRSTNFRWGGAFDLRLFFETNLCHQISFYRRELFDGIGPNLYRYRWADWDFNIR 198

CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 275 AA;

Query Match 56.2%; Score 788; DB 6; Length 275;
 Best Local Similarity 60.9%; Pred. No. 7.8e-77;
 Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;
 QY 1 MTAPEVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSGTDRTLDIANSFRPELGS 60
 DB 20 MAAPFWSIIITLNVAAVLPACLDSIAQTGDFELVLVDGSGTDETLDIANIPAPNLGE 79
 QY 61 RLNVHSGPDGPDYDAMNRGVGATGEWVFLFGADDTLYEPTTLAQVAAFLGDHAAHLVY 120
 DB 80 RLIIHRDTDQGYDAMNRGVGLATGTWLLFLGADDSLYEADTLARVAAFIGEHEPSDLVY 139
 QY 121 GVVVVRSTKSRHAGPDLDRLLFETNLCHQSIIFYRRELFDGIGPYNLYRWADWDFNIR 180
 DB 140 GDVIMRSTNFRWGGAFDLRLFKRNICHQAIIFYRGLFGTIGPYNLYRWADWDFNIR 199
 QY 181 CFSNPALITRYMDVVISBYNDMTGFSMRQGTDKFPRKRLPMYFWVAGWETCRLMLAFLKD 240
 DB 200 CFSNPALITRYMVAASYNFEGGLS-NTIYDKEFLKRLPM----- 239
 QY 241 KENRRRLALTRLIRVK 256
 DB 240 --STRLGIRLVILVR 253

RESULT 7
 AAW88309
 ID AAW88309 standard; protein; 297 AA.
 XX
 AC AAW88309;

DT 26-APR-1999 (first entry)

DE E. coli colitose or glucose transferase.

KW O antigen; O111 antigen; wbdL gene; colitose transferase;
 KW glucose transferase; diarrhoea; haemorrhagic colitis; diagnosis.

OS *Escherichia coli*.

XX WO9850531-A1.

XX 12-NOV-1998.

XX 01-MAY-1998; 98WO-AU000315.

XX 01-MAY-1997; 97AU-00006545.

XX 22-JUL-1997; 97AU-00008162.

XX (UNSY) UNIV SYDNEY.

PI Reeves PR, Wang L;
 XX WPI; 1999-059669/05.
 DR N-PSDB; AAX06748.
 XX
 PT Nucleic acid molecules specific for bacterial polysaccharide antigens -
 PT useful for detecting specific strains in, e.g. food, faeces or patient
 XX samples.
 PS Disclosure; Fig 7; 165pp; English.
 XX
 CC This is the amino acid sequence of the protein encoded by the wbdL gene
 CC of a gene cluster (see AAX06748) involved in the biosynthesis of the
 CC *Escherichia coli* O111 antigen. The protein shows low homology with *Leid*
 CC of *Neisseria gonorrhoeae*, and is predicted to be a colitose or glucose
 CC transferase. The use of nucleic acid molecules derived from particular
 CC assembly and transport genes, particularly wbd (transferase), wzx
 CC (flippase) and wzy (polymerase) genes, within O antigen gene clusters
 CC improves the specificity of methods for the detection and identification
 CC of O antigens, e.g. in testing food- or faecal-derived samples, or
 CC samples from patients. The O antigen is a major virulence factor of
 CC enteropathogenic *E. coli* strains that cause diarrhoea and haemorrhagic
 CC colitis
 XX
 SQ Sequence 297 AA;

Query Match 21.3%; Score 298; DB 2; Length 297;
 Best Local Similarity 27.7%; Pred. No. 2e-23;
 Matches 78; Conservative 52; Mismatches 86; Indels 66; Gaps 8;
 QY 1 MTAPEVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSGTDRTLDIANSFRPELGS 60
 DB 15 LDAFLVSIITATYNSLSDIAKLSQVTSQYKNTIIMDGGSSDKTLIAKSPKDD--- 71
 QY 61 RLNVHSGPDGPDYDAMNRGVGATGEWVFLFGADDTLYEPTTLAQVAAFLGDHAAHLVY 120
 DB 72 RIKIVSEKDRGIYDAMNKAVDLSIGDWVAFITGSDVVYHTDAIASLAKGVMVSNAPVY 131
 QY 121 GVVVVRSTKSRHAGP-----FDLDRLLFETNLCHQSI-----FYREL 158
 DB 132 G-----RTAHEGPDNRNIGFSGSEWNLGTFKFNYYKCNLPDIPMSAIYSRDFRNER 184
 QY 159 FDGIGPYNLYRWADWDFNIRCF-----SNPALITRYMDVVISBYNDMTGFSMRQGT 211
 DB 185 FD-----IKLIVADADWFLRCFKWSKEKSPYFINDTTFIVRMGYGGVSTDISSQVKT 238
 QY 212 DKFPRKRLPMYFWVAGWETCRLMLAFLKDKENRRRLALTRLI 253
 DB 239 TLE-----SFTVRKKNNISCLNIQLI 259

RESULT 8

ABU50169
 ID ABU50169 standard; protein; 247 AA.
 XX
 AC ABU50169;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #35696.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX *Yersinia pestis*.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.


```

Qy      176 DFNIRCSNPALITRYMDV--VISEYNDWTG 204
          :|:|         |||::: |:|:| |
Db      179 EFLIRA-----ALVCEPVTIRVCLECF-DTGG 204

RESULT 11
ABP26805
ID   ABP26805 standard; protein; 321 AA.
XX
XX   ABP26805;
XX
DT   02-JUL-2002 (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 2786.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus
KW   group A streptococcus; Streptococcus pyogenes
KW   antiinflammatory; infection; vaccine; meningi
XX   OS   Streptococcus agalactiae.
XX
XX   WO200234771-A2.
XX
XX   02-MAY-2002.
XX
XX   29-OCT-2001; 2001WO-GB004789.
XX
XX   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
XX   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
XX   Tettelin J, Massignani V, Margarit Y Rosi, C
PI   Tettelin H;
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN67436.
XX
XX   New Streptococcus protein for the treatment o
PT   disease caused by Streptococcus bacteria, su
PT   detecting a compound that binds to the proteo
PS   Claim 1; Page 3429; 4525pp; English.
XX
XX   The invention relates to a protein (ABP25413)
CC   streptococcus/GBS (Streptococcus agalactiae)
CC   (Streptococcus pyogenes), comprising one of s
CC   the specification. The proteins have antibac
CC   activity. (I), nucleic acids encoding (I), AS
CC   antibodies that bind (I) are used in the manu
CC   the treatment or prevention of infection or c
CC   Streptococcus bacteria, particularly S. aga
CC   Nucleic acids encoding (I) are used to detecte
CC   biological sample. (I) is used to determine v
CC   (I). A composition comprising (I) or a nucleo
CC   used as a vaccine or diagnostic composition.
CC   Streptococcus that is prevented or treated m
CC   acid encoding (I) may be used to recombinantl
CC   used in gene therapy. Antibodies to (I) are u
CC   chromatography, immunoassays, and distinguis
CC   Streptococcus proteins
XX
XX   Sequence 321 AA;
SQ

Query Match           14.4%; Score 201.5; DR
Best Local Similarity 30.1%; Pred. No. 7.4e-1
Matches 52; Conservative 33; Mismatches

Qy      7 SIIPFTNAAVTLQACLSIVGQTREYVEVLVDGCG
          |||||::|       |:|:| |

```

D b	5	SIIPVYNVQSFNLCIESVLQAOTYSNULEIILVNDGSTDNSGDICD-YXSEIDGRIFVPH	63
Q y	67	GPDDGYDAMNRGVGATGEWWFLIGADDTLYEPTTLAQVAAFGLGDHAAHSLVYG----	121
D b	64	KNNGLSDARNYG1GRATGDVIYLLDSDDYLYKEDAFIERMVEP-SEKYNSEIYLCGYVEK	122
Q y	122	-----DVVMRSTKSRHAGP-----FDRLRLFTETNLCHOSIFYRRELF	159
D b	123	REOHINIVLEDEMLETISPVQAOINIXYNDAYRAIF--TVAHMKL--KYKELF	172

RESULT 12
ABU23911
ID ABU23911 standard; protein; 336 AA.

CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)
XX
SQ Sequence 336 AA;

RESULT 13
AAW88314
ID AAW88314 standard; protein; 248 AA.

Nucleic acid molecules specific for bacterial polysaccharide antigens -
useful for detecting specific strains in, e.g. food, faeces or patient
samples.

Disclosure; Fig 8; 165pp; English.

SQ Sequence 248 AA;

DB 4 SVITVTYNNAEGLEKILUSSLSILKIKPFELIIVDGGSIDGINRVISKFI---SMNIITHV 59

Db 60 YEKDEGIYDAMNKGRLAKGDLIHYLNAGDSV-----IGD-----IYKNI-- 99

100 KEPCIKVGLFENDKILGFSSITHSNTGYCHOGVIFPKNHSE----YDLRYKICADYKLI 155

UY 235 LAFLUREN 243
: | | : |

AD16/15 / Standard; protein; 331 AA.

DT 22-APR-2004 (first entry)

XX Lactic acid bacteria: vaccine: bacterial infection: microbe.

KW flavour; aroma; texture; nutritional value; food; microorganism;

OS Lactobacillus rhamnosus; strain HN001.
XX

PD 15-JAN-2004.

XX
PP 09-211C-1999. 9911S-0147852D

PR 01-SEP-1999; 99US-0152032P.

XX
PA (GENE-) GENETICS PPS & NEW CORP LTD

XX
XX

XX
DN
New: no longer from Rochester [UN00] still useful for

PT flavor, aroma or nutritional benefits of a bioactive or probiotic
PT supplement product.

PS Claim 21; SEQ ID NO 152; 54pp; English.

The present invention relates to the isolation of novel lactic acid bacteria (*Lactobacillus rhamnosus* strain HN001) polynucleotide sequences, and the proteins encoded by them. Also disclosed are oligonucleotide probes and primers, and genetic constructs comprising the polynucleotide sequences of the invention. The polynucleotide sequences are useful for preparing a vaccine against bacterial infections or for improving the properties of microbes used in the manufacture of milk-derived products, food products, food additives, nutritional supplements, bioactive substances or probiotic supplements, and for modifying the flavour, aroma, texture and/or nutritional value of foods. They are also useful for identifying microorganisms having a trait associated with the polynucleotide. The present sequence represents a novel *L. rhamnosus* polypeptide sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification. The complete sequence data for this patent was obtained in electronic format directly from the USPTO web site at seqdata.uspto.gov.

Sequence 331 AA;

Query Match	13.5%; Score 189.5; DB 8; Length 331;
Best Local Similarity	28.2%; Pred. No. 1.6e-11;
Matches	62; Conservative 38; Mismatches 93; Indels 27; Gaps 9;
QY	4 PVFSIIPTFNAAVTLQAQLCSIGVQTREVVEVLVDGSGTDRITLDIANSRPELGLSLV 63
DB	7 PLVSIIPIYNVEKYLQRCDLSLPAQIVNIEIVLDDGSADRSLSICK-NAQODEVR 65
QY	64 VHSPPDGPVDMMRGVGATGEWFLFGADDTLYEPTTLAQVAAPL-GDHAASHLVYGD 122
DB	66 VFSTKGGVADTRNFYSVARGOWISFVDPDYV-DPDYIELYXGILVKSNATMSICQHR 124
QY	123 VMWRSTK--SRHAGPFDLRLLPETNLCHOSIFVRRELPGIGPYNLRYRVWAD---W 175
DB	125 NVYKNGKIQTNLYEGPAVL-----SHTAV--KRLYYDD----QIDTSVNAKLIPAW 170
QY	176 DFNIRCFNSPALITRYMDVWISYNDMTGFSMRQGTDFKEF 215
DB	171 VFKKIHFPKGRLL-----FDIAATYKTFFLASDSGIavgseaky 207

Search completed: February 27, 2005, 21:07:50
Job time : 169 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 19:38:19 ; Search time 3386 Seconds
(without alignments)
9004.572 Million cell updates/sec

Title: US-10-805-311-23

Perfect score: 801
Sequence: 1 atgactgcccagtggtcttc.....aacgaagcgacgaaccgtag 801

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.4	6.2	1266	BZ558529	pa98401.2
C 2	48	6.0	751	CF867014	trico009xd
C 3	48	6.0	805	CB897020	trico009xd
C 4	46	5.7	1212	BZ560978	pac82-164
C 5	45.6	5.7	660	BZ891972	hm11_0147
C 6	44.8	5.6	558	CK308368	hzma0_005
C 7	44.4	5.5	548	CK146970	ZM8B500
C 8	44.2	5.5	833	CK195343	FGAS00378
C 9	43.4	5.4	491	CA026025	HZ53N24r
C 10	43.4	5.4	658	BI960016	HVSMen002
C 11	43.4	5.4	669	BI959554	HVSMen002
C 12	43.2	5.4	608	AJ274182	AJ274182
C 13	43	5.4	802	CK201078	FGAS00959
C 14	42.4	5.3	464	CA025392	HZ52A13r
C 15	42.4	5.3	553	CA023059	HZ45B23r
C 16	42	5.2	516	AA835220	ak65a10.8
C 17	41.8	5.2	631	BI958865	HVSMen001
C 18	41.6	5.2	963	CF885491	trico082xi
C 19	41.2	5.1	1026	BE039820	OC08F05 O
C 20	40.8	5.1	399	AV423282	AV423282
C 21	40.8	5.1	445	BZ337760	ia90g07_b
C 22	40.8	5.1	661	CA092574	SCVFAM206
C 23	40.8	5.1	743	COS30618	3530_1_20
C 24	40.8	5.1	868	BZ404229	OGABG507C

C 25	40.8	5.1	920	9	CG445414	CG445414
C 26	40.8	5.1	936	9	CG337316	OGXCS52TV
C 27	40.8	5.1	951	9	CG337305	OGXCS52TH
C 28	40.6	5.1	688	2	BF430504	OG01F05T3
C 29	40.6	5.1	759	6	CB643463	OSJNEB04C
C 30	40.6	5.1	779	6	CB642703	OSJNEB03A
C 31	40.6	5.1	897	2	BE040583	OF03D08 O
C 32	40.4	5.0	822	8	CG329310	OGWLD41TV
C 33	40.4	5.0	832	8	CG330216	OGRAL86TV
C 34	40.4	5.0	832	9	CG443515	OGVGM64TH
C 35	40.4	5.0	925	9	CHS0031P	AL053013
C 36	40.4	5.0	1029	9	CL976984	DR080PH1
C 37	40.2	5.0	696	6	CD913580	OSIFCC043
C 38	40.2	5.0	710	6	CA182349	GS50.118G
C 39	40	5.0	509	4	BM377275	SCBGS7311
C 40	40	5.0	735	7	CF875058	EBem05_SQ
C 41	40	5.0	789	6	CB904315	trico37xn
C 42	40	5.0	835	7	CK195936	trico37xn
C 43	39.8	5.0	463	2	BE608463	FGAS00438
C 44	39.8	5.0	505	4	BI947516	sq33e11.y
C 45	39.8	5.0	642	2	BE969846	HVSMEL000
						BE969846

ALIGNMENTS

RESULT 1
BZ558529/c 1266 bp DNA linear GSS 17-DEC-2002
LOCUS pa98401.220.xl pac82-164 Pseudomonas aeruginosa genomic clone
DEFINITION pa98401.220, genomic survey sequence.
ACCESSION BZ558529
VERSION BZ558529.1 GI:27173107
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES

source
Location/Qualifiers
1..1266
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.220"
/clone_lib="pac82-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 6.2%; Score 49.4; DB 8; Length 1266;
Best Local Similarity 57.1%; Pred. No. 0.0076;
Matches 109; Conservative 0; Mismatches 81; Indels 1; Gaps 1;
QY 4 ACTGGCCAGTGTTCCTCGATATTCCTACCTTCATGCGCGGTGACGCTGCAAGCC 63
Db .259 ACGAGCCATGGTTTCGTGATCATCGTCTTCCTATACCGAGAAAGTACATAGAGG 200
QY 64 TGCCTCGAAGCATCGTCGGGACACCTACCGGAAGTGAAGTGGTCTCTTGTGACGCG 123

Db 199 ACATCGCCAGCGTGTC-AGCAGCATACCCGACGTTGGAATCTGCTGGTGTGACGAT 141
 QY 124 GGTTCGACGATCGGACCTCGACATCGCGAACAGTTTCGCCGCCGAACTCGGCTCGCGA 183
 Db 140 GGTTCGACGATCGGACGCTCGAGCTTTTTCGCGGCTCCAGGCGAAGTACGCTTCGAC 81
 QY 184 CTGGTCGTTC 194
 Db 80 CTCAGGTCCTCA 70

RESULT 2
 CF867014 751 bp mRNA linear EST 31-OCT-2003
 LOCUS
 DEFINITION
 Hypocrea jecorina cDNA clone trico09xd01, Version 6 October 2003
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 751)
 AUTHORS
 Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
 Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
 Ward, M. and Dean, R.A.
 TITLE
 Characterization of the protein processing and secretion pathways
 in a comprehensive set of expressed sequence tags from Trichoderma
 reesei
 JOURNAL
 FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
 COMMENT
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph_dean@ncsu.edu
 Seq primer: LT-F1 primer.
 Location/Qualifiers
 1..751
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico09xd01"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October
 2003"
 /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN
 Query Match 6.0%; Score 48; DB 7; Length 751;
 Best Local Similarity 54.0%; Pred. No. 0.018;
 Matches 121; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 QY 74 GCATCGTCGGGACGACCTACCGGAGTGGAGTGGTCTTGTGACGCGGTTGACCG 133
 Db 206 GCATTTTCACAGACACCACTCGATTGGAACTCATCATCTGTCAGACGGCTCGCCG 265
 QY 134 ATCGGACCTCGACATCGCGAACAGTTTCGCCCGGAACTCGGCTCGGACTGGTCTGTTTC 193
 Db 266 ACGGACCCCAAGAGTGGCAACACGCTCGTCAAGGCTTACGCCCCCAACGTCGTCCTCA 325
 QY 194 ACAGCGGGCCGATGAT---GGCCCTACGACCCATGAACCGCGGCTCGGCGTAGCCA 250
 Db 326 AGACTCGCTCCGCAAGCTGGGCGCTCGGCACCGCCTACGTCACGCGCTCGCAGTTCGTC 385
 QY 251 CAGGCGAATGGGTACTTTT---TAGGCGCGGACGACACCCCTCTAC 294

Db 386 CGGCAACTTCGTATCATCATGAGCGGACTTCAGCCACCAC 429
 RESULT 3
 CB897020 805 bp mRNA linear EST 02-JUL-2003
 LOCUS
 DEFINITION
 trico09xd01 T.reesei mycelial culture, Version 3 april Hypocrea
 jecorina cDNA clone trico09xd01, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 805)
 AUTHORS
 Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
 Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
 Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
 Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
 TITLE
 Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus Trichoderma reesei
 J. Biol. Chem. 278 (34), 31988-31997 (2003)
 MEDLINE
 PUBMED
 COMMENT
 Contact: Pamela K. Foreman
 Genencor Intl.
 925 Page Mill Road, Palo Alto, CA 94304, USA
 Tel: (650) 846-7635
 Fax: (650) 621-7817
 Email: Pforeman@genencor.com
 Seq primer: LT-F1 primer.
 Location/Qualifiers
 1..805
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico09xd01"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 3 april"
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN
 Query Match 6.0%; Score 48; DB 6; Length 805;
 Best Local Similarity 54.0%; Pred. No. 0.018;
 Matches 121; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 QY 74 GCATCGTCGGGACGACCTACCGGAGTGGAGTGGTCTTGTGACGCGGTTGACCG 133
 Db 260 GCATTTTCACAGACACCACTCGATTGGGAACTCATCATCTGTCAGACGGCTCGCCG 319
 QY 134 ATCGGACCTCGACATCGGAAACAGTTTCGCCCGGAACTCGGCTCGGACTGGTCTGTTTC 193
 Db 320 ACGGACCCCAAGACTCGCCACACAGCTCGTCAAGGCTTACGCCCCCAACGTCGTCCTCA 379
 QY 194 ACAGCGGGCCGATGAT---GGCCCTACGACCCATGAACCGCGGCTCGGCGTAGCCA 250
 Db 380 AGACTCGCTCCGCAAGCTGGGCGCTTCGGCACCGCCTACGTCACGCGCTCGCAGTTCGTC 439
 QY 251 CAGGCGAATGGGTACTTTT---TAGGCGCGGACGACACCCCTCTAC 294
 Db 440 CGGCAACTTCGTATCATCATGAGCGGCTTCAGCCACCAC 483

RESULT 4
 BZ560978/c 1212 bp DNA linear GSS 17-DEC-2002
 LOCUS
 DEFINITION
 pacs2-164_2973.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
 pacs2-164_2973, genomic survey sequence.
 ACCESSION
 BZ560978

VERSION	B2560978.1	GI:27180039	
KEYWORDS	GSS:		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS	Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burris, J.L., Kaul, R. and Olsen, M.V.		
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library		
JOURNAL	J. Bacteriol. (2002) In press		
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 20622216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: shotgun.		
FEATURES	source		
	1. .1212		
	/organism="Pseudomonas aeruginosa"		
	/mol_type="genomic DNA"		
	/strain="2-164"		
	/db_xref="taxon:287"		
	/clones="pacs2-164 2973"		
	/clone_lib="pacs2-164"		
	/notes="clinical isolate 2-164 Whole genomic shotgun library."		
ORIGIN			
Query Match	5.7%;	Score 46; DB 8; Length 1212;	
Best Local Similarity	48.8%;	Pred. No. 0.071;	
Matches	124; Conservative	0; Mismatches 130; Indels 0; Gaps 0;	
QY	100	GTGGAAGTGTGCTTGTTCGACGGCGGTTCGACCGATCGGACCTCGACATCGCGAACAGT 159	
Db	256	GTGCGTCTGGCCCTTTTCGATCTCGATAACACCCCTGCTCGCGCGGACGCGACCATAGC 197	
QY	160	TTTCGCGCCGGAATCGGCTCGGACTCGGTTCACAGCGGCGCGATGATGGCCCTTAC 219	
Db	196	TGGGCGGAATGGCTGTGCCAGCGGCGCTGTGTCGACGCGCGGAATACCGAGCGCGCAAC 137	
QY	220	GAGCCCATGAACCGCGCGTTCGGGTAGCCACAGCGGAATGGGTACTTTTATAGGCGCC 279	
Db	136	GAGCCTTCTATGCCGACTAGTTCGCGGAGGCTCGAAGTCTCTGCTACGAGCCTTC 77	
QY	280	GAGCAGACCTCTTACGAAACCAACACATGTGGCCAGGTAGCGGCTTTCTTCGGCGACAT 339	
Db	76	ACCAGCGCATCTCGGTTCGACCGAGATGGCGCGGGATCTCTAGAGTCGACCTGCA 17	
QY	340	CGCGCAGCCATCT 353	
Db	16	GAGGCGAGCTCT 3	
RESULT 5			
BZ891972			
LOCUS	Hm11.0147.x1 049.ab1	660 bp	DNA linear
DEFINITION	43049 genomic 5', genomic survey sequence.		GSS 30-JUL-2003
ACCESSION	BZ891972		Library Haloarcula marismortui ATCC
VERSION	BZ891972.1		
KEYWORDS	GSS.		
ORGANISM	Haloarcula marismortui ATCC 43049		
	Haloarcula marismortui ATCC 43049		
	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;		
	Halobacteriaceae; Haloarcula.		
REFERENCE	1 (bases 1 to 660)		
AUTHORS	Goo, Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M., DasSarma, S., Ng, W.W. and Hood, L.		
TITLE	Low-pass Sequencing for Microbial Comparative Genomics		

```

/organism="Hypocrea lixi"
/mol_type="mRNA"
/db_xref="taxon:5544"
/dev_stage="mycelium"
/lab_host="E.coliDH10"
/clone_lib="Trichoderma harzianum mycelium cDNA library"
/note="Vector: pBLUESCRIPT II SK+"

ORIGIN
Query Match          5.6%; Score 44.8; DB 7; Length 558;
Best Local Similarity 53.1%; Pred. No. 0.14;
Matches 119; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 74 GCATCGTCGGGAGACCTACCGGAGTGGTCTTCGACGCGGTCGACCG 133
Db 128 GCACCTTTACAGAGAACCAACTGGATTTGGAACTCATCATCGTCACGACGCTCACCG 187

QY 134 ATCGACCCCTCGACATCGGAACTTTCCGCGCGAACTCGGCTCGGACTGTCGTTTC 193
Db 188 ATGGACACAGACGTCGGCAACACAGCTCGTCAAGCCCTACGCCCCCACTCGTCTCTCA 247

QY 194 ACAGCGGGCCGATGAT---GCCCCCTACGACGCGCATGAACCGCGGCTCGCGGTAGCCA 250
Db 248 AGACTCGATCCGCGCAAGCTGGGCTCGGAAACGCTACGTCACGCGCTGCACTTTGTCA 307

QY 251 CAGCGCAATGGTACTTTTTTAGGCGCGGACGACACCTCTAC 294
Db 308 CGGGCAACTTTGTCTATCATCATGAGCGCGGACTTCAGCCACAC 351

```

```

RESULT 7
LOCUS CCL146970/c
DEFINITION ZMMBBb0009F05.r ZMMBBb Zea mays genomic clone ZMMBBb0009F05 3',
genomic survey sequence.
ACCESSION CCL146970
VERSION CCL146970.1 GI:30092163
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
AUTHORS 1 (bases 1 to 548)
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
and Wing, R.
TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: T7
BACKWARD: M13r
Plate: 0009 row: F column: 05
Seq primer: M13r
Class: BAC ends.

```

```

FEATURES
source
1..548
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBb0009F05"
/lab_host="DH10B"
/clone_lib="ZMMBBb"
/note="Vector: pBelOBAC11; Site 1: HindIII; Site 2:
HindIII; Zea mays L. ssp. mays"

```

```

ORIGIN
Query Match          5.5%; Score 44.4; DB 8; Length 548;
Best Local Similarity 47.2%; Pred. No. 0.18;
Matches 135; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 46 GCGGTGACGCTCGAAGCCTCGCTCGAAGCATCGTCGGGACACCTACCGGAAGTGGA 105
Db 432 GCGCTAACGCGCGGCTGGGTTCGGGCGCTCAGGCGCCAGGGCTCCACCGACGGCGG 373

QY 106 GTGTCCTTTGTCGACGCGGTTTCGACCATCGGACCCCTCGACATCGGAAACAGTTTCGCG 165
Db 372 GTGCGCTTCGCGTTCGGCGAGCGGACTTGACAAAGTCGAGGAGCGGTGCTCACCTC 313

QY 166 CGGAACTCGGCTCGGACTGCTGTTTCACAGCGGCGCGATGATGGCCCTTACGACGCC 225
Db 312 CTTGAGTAGGTCTGACGCGCTCGTTCCTCCACGAGGCGCGCGCCCTCCCGC 253

QY 226 ATGAACCGCGCTCGGCGTAGCCACAGCGCAATGGGTACTTTTTTATAGGCGCCGACGAC 285
Db 252 AGCGCCACAGCGTTCGAAGGCTCGGTCTCGATGCGCGCGCGCGGCTCGCGTTCGCG 193

QY 286 ACCCTCTACGAAACCAACCATCGTTGCCCGAGGTAGCCGCTTTTCTCG 331
Db 192 GACCGGACGCGCCATAGCGCTTGCAAGGACGCGTGTCTCCCGCG 147

RESULT 8
LOCUS CK195343
DEFINITION FGAS003782 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
aestivum cDNA, mRNA sequence.
ACCESSION CK195343
VERSION CK195343.1 GI:39557733
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
AUTHORS 1 (bases 1 to 833)
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroché, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
Penniket, C., Roach, J.L. and Sarhan, F.
Functional Genomics of Abiotic Stress In Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.est@usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [102,703].
Plate: L3C104 row: H column: 14.
Location/Qualifiers
1..833
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
/note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
control, cold-acclimated and salt stressed wheat cultivar
Norstar. 7 mRNA populations were combined before
constructing the library; 7 day non-acclimated roots, 1,
23, and 53 days cold-acclimated at 4C, and 30 minutes, 3
hours and 6 hours treated roots with 200mM NaCl.

```

Non-acclimated and cold-acclimated plants were grown in vermiculite while salt stressed plant were grown hydroponically. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 5.5%; Score 44.2; DB 7; Length 833;
Best Local Similarity 50.2%; Pred. No. 0.22;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 47 CGGTGAGCTGCGAGCTGCTCGGAGCATGTCGGGCGAGCTTACCGGGAAGTGGAG 106
DB 547 CGGTGAGCTGCGAGCTGCTCGGAGCATGTCGGGCGAGCTTACCGGGAAGTGGAG 606

QY 107 TGTGCTCTGTCGAGCGGCTTGCACCGATCGGACCTCGACATCGGACAGTTCCTCGGC 166
DB 607 CCTACGAGAGATGCGACCGGCACACATCGGAGCATCACTCATACCGTGGCC 666

QY 167 CGGAATCGGCTCGGAGCTGCTGCTTTCACAGCGGCGCCGATGATGCCCTTACGAGCCA 226
DB 667 AGAAGCTCTGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726

QY 227 TGAACCGCGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257
DB 727 TGCACCTCGGCACATCGTCTGAGCGCGCGGA 757

RESULT 9

CA026025
LOCUS
DEFINITION
HZ53N24r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ53N24
5-PRIME, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CA026025.1 GI:24303399
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 491)

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
Barley ESTs from developing seeds
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595

Email: stein@ipk-gatersleben.de
Insert Length: 491 Std Error: 0.00
Plate: 53 row: N column: 24
Seq primer: M3rev.

FEATURES

source

1. .491
Location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:279451"
/db_xref="taxon:112509"
/clone="HZ53N24"
/tissue_type="pericarp"
/dev_stage="0-7 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP (days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To

ORIGIN

Query Match 5.4%; Score 43.4; DB 6; Length 491;
Best Local Similarity 48.2%; Pred. No. 0.35;
Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 406 TTCGACCTCGACCGCTCTATTGAGAGCAATTTGTGCCCAATCGATCTTTTACCGC 465
DB 209 TTCTACTGGGCGCCCTCTCTCGCCGAGTCCAAACGCCGACGCGCTCGAGCACCAGCTC 268

QY 466 CGTGAGCTTTTCGACGCGCATCGGCCCTTACAACTCGCTACCGAGTCTGGCGGAGCTGG 525
DB 269 AAGGACCGCTCATCGCGGCGGCCATGACCGCCACTCCGCTTCTGGAGGCGGCC 328

QY 526 GACTTCAATATTGCTGCTTCTCCAAACCGCGGCTGTATTACCCGTTACATGACGCTG 585
DB 329 GAGCTCTCTGCTTCAACTCTCTCTGCTGATGACCGGCGACAAGATCCAGATCTCTG 388

QY 586 ATTTCCGAATACAACGACATGACCGGCTTCAGCATGACGCGGAGCTGATTAAGAGTTTC 645
DB 389 AGGGGCGCGCACACGACATGAGCAAGGACATCGTGAGATGGGGCGGAGGCGGTAC 448

QY 646 AGAAAAACGGCTGC 658
DB 449 AGGCTGCTGCTGC 461

RESULT 10

BI960016
LOCUS

DEFINITION
HVSME0022N21f Hordeum vulgare subsp. vulgare cDNA clone HVSME0022N21f,
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BI960016
BI960016.1 GI:16311271
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 658)

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,
Simmons, J., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex rachis cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 546
Seq primer: AATTAACCTCACTAAAGG
High quality sequence stop: 641.
Location/Qualifiers

1. 658
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSME0022N21f"
/tissue_type="Rachis"
/lab_host="TJG121"
/clone_lib="Hordeum vulgare rachis EST library HVCNDA0015"

FEATURES
source

excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, Sali, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 900 bp"

(normal)"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders/Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 5.4%; Score 43.4; DB 4; Length 658;
 Best Local Similarity 48.2%; Pred. No. 0.36;
 Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
 QY 406 TTCGACCTCGACCGCTCTATTTCGAGACGAATTTGTGCCCAATCGATCTTTTACCGC 465
 Db 193 TTCTACTGGCGCGCTCTCCCGAGTCCAAAGCGGAGCGGCTCGAGCACAGCTC 252
 QY 466 CQTGAGCTTTTCGACGGCATCGCCCTTTACACCTCGGCTACCGGCTCTGGCGGAGCTGG 525
 Db 253 AACGACCGCTCATCGGGCGGCCCATGAGACCGCCACTCCGCTTCTGGAAGGCGCC 312
 QY 526 GACTTCAATATTCGTGCTTCTCCACCGCGGCTGATTACCGCTACATGACGCTGTG 585
 Db 313 GAGTCTCTGCTTCAACTCTCTACCTCTGTGTGATGACCGGCGACAAGATCCAGATCTCTG 372
 QY 586 ATTTCCGAATACAGCATGACCGCTTTCAGCATGAGCGAGGAGCTGATAAAGAGTTC 645
 Db 373 AGGGCGCGCACACACATGAGACGATCTGTGGAGATGGGGCGGAGGCGGTAC 432
 QY 646 AGAAACCGCTGC 658
 Db 433 AGGCTGGTCTGC 445

RESULT 11
 B1959554
 LOCUS B1959554.1 669 bp mRNA linear EST 22-OCT-2001
 DEFINITION HVSMEN020C17f Hordeum vulgare rachis EST library HVCNDA0015 (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN020C17f, mRNA sequence.
 ACCESSION B1959554
 VERSION B1959554.1 GI:16310809
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 669)
 AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D., Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R., and Main, D.
 TITLE Development of a genetically anchored and physically anchored EST resource for barley genomics: Morex rachis cDNA library
 JOURNAL Unpublished (2001)
 COMMENT Contact: Wing RA

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 418
 Seq primer: AATTAAACCTCACTAAAGGG
 High quality sequence stop: 583.
 Location/Qualifiers
 1.669
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSMEN020C17f"
 /tissue_type="Rachis"
 /lab_host="TJCl21"
 /clone_lib="Hordeum vulgare rachis EST library HVCNDA0015 (normal)"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders/Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 5.4%; Score 43.4; DB 4; Length 669;
 Best Local Similarity 48.2%; Pred. No. 0.36;
 Matches 122; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
 QY 406 TTCGACCTCGACCGCTCTATTTCGAGACGAATTTGTGCCCAATCGATCTTTTACCGC 465
 Db 130 TTCTACTGGCGCGCTCTCCCGAGTCCAAAGCGGAGCGGCTCGAGCGCCAGCTC 189
 QY 466 CQTGAGCTTTTCGACGGCATCGCCCTTTACAACTCGCTACCGGCTCTGGCGGAGCTGG 525
 Db 190 AACGACCGCTCATCGGGCGGCCCATGACCGCCACTCCGCTTCTGGAAGGCGCC 249
 QY 526 GACTTCAATATTCGTGCTTCTCCAAACCGGCGCTGATTACCCGCTACATGACGCTGTG 585
 Db 250 GAGTCTCTGCTTCAACTCTCTCTGTGGATGACCGGCGACAAGATCCAGATCTCTG 309
 QY 586 ATTTCCGAATACAGCATGACCGCTTTCAGCATGAGCGGAGCTGATAAAGAGTTC 645
 Db 310 AGGGCGCGCACACACATGAGCAAGACATCTGTGGAGATGGGGCGGAGGCGGTAC 369
 QY 646 AGAAACCGCTGC 658
 Db 370 AGGCTGGTCTGC 382

RESULT 12
 AJ274182

LOCUS AJ274182 608 bp mRNA linear EST 15-NOV-1999
 DEFINITION AJ274182 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
 cDNA clone Ma#1818, mRNA sequence.
 ACCESSION AJ274182
 VERSION AJ274182.1 GI:64333555
 KEYWORDS EST.
 SOURCE Metarhizium anisopliae
 ORGANISM Metarhizium anisopliae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
 Clavicipitaceae; Metarhizium.
 1 (bases 1 to 608)
 REFERENCE Screen, S.E., Mathur, P. and St. Leger, R.J.
 AUTHORS EST analysis of the insect pathogenic fungus Metarhizium anisopliae
 TITLE Unpublished (1999)
 JOURNAL Contact: Screen SE
 COMMENT Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA.
 FEATURES
 source
 1. .608
 /organism="Metarhizium anisopliae"
 /mol_type="mRNA"
 /strain="ARSEF 2575"
 /db_xref="taxon:5530"
 /clone="Ma#1818"
 /note="Vector: Unizap; Metarhizium anisopliae was grown on
 insect cuticle for 24 hours. A cDNA library was
 constructed in the unidirectional Lambda vector, Unizap"

LOCUS CK201078 802 bp mRNA linear EST 08-DEC-2003
 DEFINITION FGAS009597 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
 aestivum cDNA, mRNA sequence.
 ACCESSION CK201078
 VERSION CK201078.1 GI:39563468
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (bases 1 to 802)
 REFERENCE Allard, P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocque, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
 Penniket, C., Roach, J.L. and Sarhan, P.
 AUTHORS Functional Genomics of Abiotic Stress in Wheat and Canola Crops
 TITLE Unpublished (2003)
 JOURNAL

LOCUS CK201078 464 bp mRNA linear EST 23-OCT-2002
 DEFINITION HZ52A13r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ52A13
 5-PRIME, mRNA sequence.
 ACCESSION CA025392
 VERSION CA025392
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 464)
 REFERENCE Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 AUTHORS Barley ESTs from developing seeds
 TITLE Unpublished (2002)
 JOURNAL Contact: Stein Nils
 COMMENT Molecular Markers Group, Department Genbank.

COMMENT Contact: Wm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas_est@cs.usask.ca
 This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region [101,763].
 Plate: L3C113 row: C column: 12.
 Location/Qualifiers
 1. .802
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
 /note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
 control, cold-acclimated and salt stressed wheat cultivar
 Norstar. 7 mRNA populations were combined before
 constructing the library; 7 day non-acclimated roots, 1,
 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3
 hours and 6 hours treated roots with 200mM NaCl.
 Non-acclimated and cold-acclimated plants were grown in
 vermiculite while salt stressed plant were grown
 hydroponically. First strand synthesis in this library was
 done in the presence of methylated dCTP thereby protecting
 from internal cleavage with NotI."

FEATURES
 source
 1. .802
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
 /note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
 control, cold-acclimated and salt stressed wheat cultivar
 Norstar. 7 mRNA populations were combined before
 constructing the library; 7 day non-acclimated roots, 1,
 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3
 hours and 6 hours treated roots with 200mM NaCl.
 Non-acclimated and cold-acclimated plants were grown in
 vermiculite while salt stressed plant were grown
 hydroponically. First strand synthesis in this library was
 done in the presence of methylated dCTP thereby protecting
 from internal cleavage with NotI."

ORIGIN
 Query Match 5.4%; Score 43; DB 7; Length 802;
 Best Local Similarity 50.2%; Pred. No. 0.49; Mismatches 106; Conservative 0; Indels 105; Gaps 0;
 Matches 106; Conservative 0; Indels 105; Gaps 0;
 QY 47 CGGTGACGCTGCAAGCCCTCGGAGCATCTCGGAGCAGACCTTACCGGGAAGTGGAG 106
 DB 253 CGGTGACGCTGCAAGCCCTCGGAGCATCTCGGAGCAGACCTTACCGGGAAGTGGAG 312
 QY 107 TGGTCTTGTGACGCGGCTTGGACCGATCGGACCTCGACATCGGGAACAGTTTCGGCC 166
 DB 313 CCTACGAGGATGCCACCGCCCAACATCGCGGAGCTCAACCTCATCACCGTCGGCC 372
 QY 167 CGGAACCTGCGCTCGGACTGCTGTTACAGCGGCGCGGATGATGCGCCCTACGACGCCA 226
 DB 373 AGAAGCTCTGGATCCCGCTGCGCTGCGAGTCCAGTCCGACCGCGCGCGCGCGTCT 432
 QY 227 TGAACCGCGGCTCGGCTAGCCACAGCGCA 257
 DB 433 TCCACCTGCCACATCTCGACGCGCGCGCGCA 463

RESULT 14
 CA025392
 LOCUS
 DEFINITION HZ52A13r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ52A13
 5-PRIME, mRNA sequence.
 ACCESSION CA025392
 VERSION CA025392
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 464)
 REFERENCE Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
 AUTHORS Barley ESTs from developing seeds
 TITLE Unpublished (2002)
 JOURNAL Contact: Stein Nils
 COMMENT Molecular Markers Group, Department Genbank.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 19:42:08 ; Search time 185 Seconds
(without alignments)
7084.636 Million cell updates/sec

Title: US-10-805-311-23

Perfect score: 801

Sequence: 1 atgactgcgcagtggtctc.....aacgaagcgcagaccgtag 801

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	346	43.2	32155	US-08-311-731A-1	Sequence 1, Appli
2	346	43.2	4403765	US-09-103-840A-2	Sequence 2, Appli
3	346	43.2	4411529	US-09-103-840A-1	Sequence 1, Appli
4	58.4	7.3	912	US-09-252-991A-12584	Sequence 12584, A
5	56.4	7.0	906	US-09-252-991A-16525	Sequence 16525, A
C 6	55.6	6.9	1060	US-09-072-596-306	Sequence 306, App
7	55.6	6.9	1060	US-09-072-596-311	Sequence 311, App
C 8	47.8	6.0	1022	US-09-072-596-325	Sequence 325, App
9	47.8	6.0	1022	US-09-072-967-330	Sequence 330, App
10	43.8	5.5	1332	US-09-902-540-4672	Sequence 4672, App
11	43.8	5.5	24754	US-09-902-540-1230	Sequence 1230, App
12	43.6	5.4	36412	US-08-311-731A-132	Sequence 132, App
13	42.8	5.3	5859	US-08-312-387B-1	Sequence 1, Appli
14	42.8	5.3	5859	US-08-312-387B-7	Sequence 7, Appli
15	42.8	5.3	5859	US-08-683-426-1	Sequence 1, Appli
16	42.8	5.3	5859	US-08-683-426-7	Sequence 7, Appli
17	42.8	5.3	5859	US-08-683-458-1	Sequence 1, Appli
18	42.8	5.3	5859	US-08-683-458-7	Sequence 7, Appli
19	42.8	5.3	5859	US-08-878-360-1	Sequence 1, Appli
20	42.8	5.3	5859	US-08-878-360-7	Sequence 7, Appli
21	42.8	5.3	5859	US-08-478-140B-1	Sequence 1, Appli
22	42.8	5.3	5859	US-09-333-412-1	GENERAL INFORMA
23	42.8	5.3	5859	US-09-333-412-7	GENERAL INFORMA
24	42.8	5.3	5859	US-09-338-943-1	Sequence 1, Appli
25	42.8	5.3	5859	US-10-007-267A-1	Sequence 1, Appli
26	42.8	5.3	5859	US-10-007-267A-7	Sequence 7, Appli
27	42.6	5.3	1431	US-09-489-039A-1316	Sequence 1316, App

C 28	42	5.2	2730	4	US-09-902-540-5705	Sequence 5705, Ap
29	42	5.2	50725	4	US-09-902-540-1271	Sequence 1271, Ap
C 30	41.8	5.2	1131	4	US-09-902-540-6687	Sequence 6687, Ap
31	41.8	5.2	9165	4	US-09-902-540-1050	Sequence 1050, Ap
C 32	41.4	5.2	1521	4	US-09-252-991A-11805	Sequence 11805, A
33	41.4	5.2	2505	4	US-09-252-991A-11574	Sequence 11574, A
C 34	41.2	5.1	1443	4	US-09-252-991A-12762	Sequence 12762, A
35	40.8	5.1	320	3	US-09-060-756-467	Sequence 467, App
C 36	40.8	5.1	320	4	US-09-670-314-467	Sequence 467, App
37	40.8	5.1	1836	4	US-09-266-965-52	Sequence 52, Appl
C 38	40.8	5.1	53500	4	US-09-266-965-76	Sequence 76, Appl
39	40.4	5.0	2811	4	US-09-266-965-28	Sequence 28, Appl
C 40	40.4	5.0	18034	4	US-09-266-965-75	Sequence 75, Appl
41	39.2	4.9	786	4	US-09-894-844-51	Sequence 51, Appl
C 42	39.2	4.9	2004	4	US-09-252-991A-11865	Sequence 11865, A
43	39.2	4.9	12732	3	US-09-060-756-1	Sequence 1, Appli
C 44	39.2	4.9	12732	4	US-09-670-314-1	Sequence 1, Appli
C 45	39.2	4.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-311-731A-1/c
; Sequence 1, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311.731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
; US-08-311-731A-1

Query Match 43.2%; Score 346; DB 4; Length 32155;
Best Local Similarity 71.0%; Pred. No. 5.7e-95;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy	1	ATGACTGCGCCAGTGTTCGATAATATCCCTTACCTTCAATGCGAGCGGTGACGCTGCAA	60
Db	15076	ATGCGCGCAACAATGTTTTGCATCATCATCCACCTTGAACGTGTGCTGCGGTATTGCGT	15017
Qy	61	GCCTGCTCGGAAGCATCGTGGCGAGACCTACCGGGAAGTGGAAAGTGGTCTCTTCGAC	120
Db	15016	GCCTGCTCGACAGCATCGCCCGTCAAGCTCGGGTGACTTCGAGCTGGTACTTGGTTCGAC	14957
Qy	121	GGCGGTTTCGACCGATCGGACCCCTCGACATCGGGAAACAGTTTTCCGCCCGGAACTCGGCTCG	180
Db	14956	GGCGGCTCGAGCGACGAAACCCCTCGACATCGCAACATTTTTCGCCCCCAACCTCGCGCAG	14897
Qy	181	CGACTGGTCTGTTACAGCGGGGCCGATGATGGCCCTTACGACGCATGAACCGCGGGCTC	240
Db	14896	CGGTTGATCATTTATCGGCACACCGACCGAGGGCGTCTACGAGCCATGAACCGCGCGCTG	14837
Qy	241	GGCGTAGCCACAGCGGAATGGGTACTTTTTTAGCGCGCGAGCACACCTCTACGAACA	300
Db	14836	GACTCGCCACCGGAACGTGGTTGCTCTTTCTGGGCGCGAGCAGACCTGTACGAGGCT	14777
Qy	301	ACCAGCTTGCCCGAGGTAGCGGCTTTTCTCGGCGACCATCGCGCAAGCCATCTTGTCTAT	360
Db	14776	GACACCTGSGCGSGGTGGCGGCTTCATTTGCGRAACAGCGCCAGCATCTGTATAT	14717
Qy	361	GGCGATGTTGTATGGTTTCGAGAAAGCGGCATGCGCGACCTTTTCGACCTCGACGCG	420
Db	14716	GGCGAGCGTATCATGGGCTCAACCAATTTCCGTGGGGTGGGCGCTTCGACCTCGACGCT	14657
Qy	421	CTCTATTTTCAGACGAATTTGCGCAACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480
Db	14656	CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTCTACGCGCGGACTCTTCGCG	14597
Qy	481	GGCATCGGCCCTTACAACTCGCGCTACCGAGTCTGGCGGAGCTGGGACTTCAATATTCGC	540
Db	14596	ACCATCGGTCCCTACAACTCCGCTACCGGGTCTTGGCGCATGCGGACTTCAATATTCGC	14537
Qy	541	TGCTTCTCGAACCGGGGCTGATTACCGGTACATGGACGTCGTGATTTCCGAATACAAAC	600
Db	14536	TGCTTTTCCAAACCGCGCTCGTCAACCGGTACATGCACTGTCGTGGCAAGCTACAAAC	14477
Qy	601	GACATGACCGGCTTCAGCATAGGCGAGGGGACTGTATAAGAGTTTCAGAAAAACGGCTGGCA	660
Db	14476	GAATTCGGCGGCTCAGCAATACGATCG--TCGACAGGAGTTTTTGNAGCGGCTGCG	14420
Qy	661	ATGTAC	666
Db	14419	ATGTCC	14414

RESIT.T 2

```

RESOL 2
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequ

```

; OTHER INFORMATION: represent a, t, c or g									
US-09-103-840A-2									
Query Match	43.2%	Score 346;	DB 3;	Length 4403765;					
Best Local Similarity	71.0%	Pred. No. 5.6e-94;							
Matches 473;	Conservative	0;	Mismatches 190;	Indels 3;	Gaps 1;				
Qy	1	ATGACTGCGCGAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAAA	60						
Db	3303798	ATGCGCGACCAATGTTTTCGATCATCATCCACACTTGAACTGGCTGCGGTATTGCGCT	3303857						
Qy	61	GCCTGCCCTCGAAGCATCGTGGGCGACACTACGGGAAGTGAAATGGTCTTTGTTCGCAC	120						
Db	3303858	GCCTGCCCTCGACAGCATCGCCCGTCAGACTCGCGGTGACTTCGAGCTGGTACTGGTTCGAC	3303917						
Qy	121	GGGGTTTCGACCCGATCGGACCCCTCGACATCGCGAAACAGTTTTCGCCCGGAACCTCGGCTCG	180						
Db	3303918	GGCGGCTCGACGACGAGAAACCTTCGACATCGCCAAATTTTCGCCCCCAACCTCGGGCGAG	3303977						
Qy	181	CGACTGGTCTGTTCACAGCGGCGCCGATGATGGCCCTTACGACGCCATGAACCGCGGCGTC	240						
Db	3303978	CGGTTGATCATTTTCATCGGACACCGACCGAGGCGTCTACGACGCCATGAACCGCGGCGTG	3304037						
Qy	241	GGCGTAGCCACAGCGGAATGGGTACTTTTTTATAGGCGCGGACGACACCTCTACGAACCA	300						
Db	3304038	GACCTGGCCACCGGAACGTTGGTTCCTTCTGGCGCGGACGACGACCTGTACGAGGCT	3304097						
Qy	301	ACCACGTTGGCCACGAGTAGCGGCTTTTCTCGGCGACCATGCGGCAAGCCATCTTGCTAT	360						
Db	3304098	GACACCTTCGCGCGGGTGGCGGCTTCATTGGCGAAACGAGGCCAGCGATCTGGTATAT	3304157						
Qy	361	GGCGATGTTGTGATCGCTTCGAGAAAGCGGCGATGCGGACCTTTTCGACCTTCGACCCG	420						
Db	3304158	GGCGAGTGATCATGGCTCAACCAATTTCCGCTGGGGTGGCGCTTCGACCTTCGACCGT	3304217						
Qy	421	CTCCTATTTCGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480						
Db	3304218	CTGTGTTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACCTCTTCGGC	3304277						
Qy	481	GGCATCGGCCCTTACACCTGGCGCTACCGAGTCTGGCGGACCTGGGACCTTCAATATTCG	540						
Db	3304278	ACCATCGGTCCTTACAACTCCGCTACCGGGTCTTCGGCGGACCTGGGACTTCAATATTCG	3304337						
Qy	541	TGCTTCTCCAAACCGGCGCTGATTACCCGCTACATGACGCTCGTGTATTTCCGAATACAAC	600						
Db	3304338	TGCTTTTCCAAACCGGCGCTCGTCAACCGCTACATGACGCTGCTGTTGCAAGCTACAC	3304397						
Qy	601	GACATGACCGGCTTCAGCATGAGCGGAGCTGTATAAGAGTTTCAGAAACCGGCTGCCA	660						
Db	3304398	GAATTGCGCGGCTCAGCAATACGATCG---TCGACAAAGGAGTTTTTTGAAGCGGCTCGCG	3304454						
Qy	661	ATGTAC	666						
Db	3304455	ATGTCC	3304460						

RESIT.T 3

```

RESOUR. 3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```

LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 43.2%; Score 346; DB 3; Length 4411529;
Best Local Similarity 71.0%; Pred. No. 5.6e-94;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGGCCAGTGTCTCGATAATATCCCTACCTCAATGACGCGGTGACGCTGCAA 60
DB 3309524 ATGCGCGACCAATGTTTCGATCATATCCCACTTGAACGTGGCTGCGGTATGCTT 3309583
QY 61 GCTTGCCTCGGAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGAAAGTGTCTCTTCGAC 120
DB 3309584 GCTTGCCTCGAGCATCGCCGTCGAGACCTGGGTGATTCGAGCTGTTCTGCTGAC 3309643
QY 121 GCGGTTCGACCGATCGACCTCGACATCGGAACAGTTTCGCGCCGGAJACTCGGCTCG 180
DB 3309644 GCGGCTCGAGCGAGAAACCTCGACATCGCAACATTTTCGCCCCCAACCTCGGCGAG 3309703
QY 181 CGACTGTGCTTCACAGCGGCGCCGATGATGSCCCTACGAGCCCATGAACCGGCGCTC 240
DB 3309704 CGTTGATCATTCATCGCGACACCGACGAGCGGCTCTACGAGCCCATGAACCGGCGCTG 3309763
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA 300
DB 3309764 GACCTGGCCACCGGAACGTGGTTGCTCTTCTGGGCGGAGACGACCTGTACGAGCT 3309823
QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGCGACCAATGCGGCAAGCCATCTTGTCTAT 360
DB 3309824 GACACCTGGCGCGGTGGCGCCTTCAATGGCGAAACAGAGCCGAGCATCTGGTATAT 3309883
QY 361 GCGGATGTTGATCGCTTCAGGAAGCGGCATCGCGACCTTCGACCTCGACCGC 420
DB 3309884 GCGGACGTGATCATCGCTCAACCAATTTCCGCTGGGTGGCGCTTCGACCTCGACCGT 3309943
QY 421 CTCCTATTGAGACGAATTTGTGCCCAACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 3309944 CTGTTGTTCAAGCGCAACATCTGCCATCAGGGATCTTACCGCGCGGACTCTTCGGC 3310003
QY 481 GGCATCGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
DB 3310004 ACCATCGGTCTCCATACACTCGCTACCGGGTCTCGGCGGACTTCAATATTGCG 3310063
QY 541 TGCTTCTCAACCCCGGCTGATACCGCTACATGAGCTTCTGATTTCCGATACAAAC 600
DB 3310064 TGCTTCTCAACCCCGGCTGATACCGCTACATGAGCTTCTGATTTCCGATACAAAC 3310123
QY 601 GACATGACCGGCTTCAGCATGAGGCGGACTGATAAGAGTTTCAGAAAACGGCTGCGCA 660
DB 3310124 GAATTCGGGGGCTCAGCAATACGATCG---TCGACAGGAGTTTTCGAGGGGCTGCGC 3310180
QY 661 ATGTAC 666
DB 3310181 ATGTCC 3310186

RESULT 4
US-09-252-991A-12584
Sequence 12584, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12584
LENGTH: 912
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12584

Query Match 7.3%; Score 58.4; DB 4; Length 912;
Best Local Similarity 49.8%; Pred. No. 1.2e-07;
Matches 208; Conservative 0; Mismatches 201; Indels 9; Gaps 2;

QY 7 GCGCAGGTCTTCGATAATATCCCTACCTCAATGACGCGGTGACGCTGCAAGCCTGCG 66
DB 58 GCGCGCTGGTATCCGTCGTAGCACCTCTTCAACGCGGAAAGTACCTGGAAGAAGCC 117
QY 67 CTCGGAGCATCGTCGGGCGAGACCTACCGGGAAGTGGAGTGTCTTGTTCGACGCGCGT 126
DB 118 CTGCGCAGCATCTACGAGCAGGACTACCCGAATTTTCGAAGTGATCATCTCGACGACGT 177
QY 127 TCGACCGATCGGACCTTCGACATCGCAACAGTTTTCGCCCGGAACTCGGCTCGGACTG 186
DB 178 TCCACCGACAACGCTAGCCATCTCGAGCAGTTTCGAGAAAGTCCACGGTTCCAGC-- 235
QY 187 GTCGTTCAAGCGGCGCGATGATGGCCCTTACGACGCTATGAACCGCGCGGTGCGGTA 246
DB 236 ---TCTACCGCAGCAGAACGAGGCGTCAGCGCAGGCTGAACTTCGGTCTGCAACAC 291
QY 247 GCGACGCGCATGGGTACTTTTTTTAGGCGCGGACGACACCTCTACGAACCAACACG 306
DB 292 GCGCGCGGCTAGTACGTCAGCCACCGGACCTGGAACGACA---TCATGCTCGCGCACTCG 348
QY 307 TTGCGCCAGTGGTGGCTTTTCTCGGACCATCGGCAAGCCATCTTGTCTATGGCGAT 366
DB 349 TTGAGCGTGGCGCGGCTGCTCGACGACCCGGAAGTGGGTGCTGGGCGCTG 408
QY 367 GTTCGTGATGCTTCGACGAAGCCGCGATGCGGACCTTTTCGACCTCGACCGGCTCC 424
DB 409 GTGATCTACATCGACAGCGAGGCGCAGGAACCAAGGCGGACGCAACCGCATCC 466

RESULT 5
US-09-252-991A-16525
Sequence 16525, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16525
LENGTH: 906
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16525

Query Match 7.0%; Score 56.4; DB 4; Length 906;
Best Local Similarity 56.5%; Pred. No. 4.9e-07;
Matches 105; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 9 GCCAGTGTCTCGATAATATCCCTACCTCAATGACGCGGTGACGCTGCAAGCCTGCT 68
DB 48 GCCATTGGTTTCCGTGATCATCGTTCTTATACCAACGAAAGTACATAGAGCGAGCAT 107
QY 69 CGGAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGTCTTGTTCGACGCGGCTTC 128

Db 108 TGCCAGCGTGGCTGCGCAGACATACCGCAAACGTGGAACTGCTGGTGGGACGATGGTTC 167
Qy 129 GACCGATCGGACCTCGACATCGGAAACAGTTTCGCCCGGAACTCGCGCTCGGACTGGT 188
Db 168 CAGCGATCGCAGCGTCGAGCTTTTTCGCGCGGCTCCAGCGGAAGTAGTACGCGCTTCGACCTCAG 227
Qy 189 GGTCA 194
Db 228 GGTCA 233

RESULT 6

US-09-072-596-306/c
; Sequence 306, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 306:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-072-596-306

Query Match 6.9%; Score 55.6; DB 3; Length 1060;
Best Local Similarity 50.2%; Pred. No. 9.2e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;
Qy 338 ATGCGGCAAGCAATCTTGTCTATGCGGATGTTGTGATCGGTTTCGACGAAAGCCGGCATG 397
Db 744 ANNMGACCCNAAKATTTGNAMWGGGAAVNTMTGSGTTC-ACCAATTTCGGGGG 686
Qy 398 CCGGACCTTCGACCTCGACCGCCCTCTATTGAGACGAATTTGTGCGCACCAATCGATCT 457
Db 685 GGGNCCTTTGTMCTTGACCTTTTGTGTTTCNAGGSAATTS-KCCATNSNGGWWTT 627
Qy 458 TTACCGCGGTGAGCTTTTTCGACGCGCATCGGCCCTTACAACTCGCGCTACCGAGTCTGGG 517

Db 626 TYACNCCCGGATTTTGGGCAMCAITGTTGCCAMCMCTTCGKAACNGGTTCTKTCG 567
Qy 518 CGGACTGGGACTTCAATATTCGCTGCTCTCCACACCGCGCGCTGATTACCGCTACATGG 577
Db 566 CCGANTGGGANTTCAATATTTGGNTGTTTTCNAACCCARG-GTTGTTNACCSGNAACATGN 508
Qy 578 A 578
Db 507 A 507

RESULT 7

US-09-072-967-311/c
; Sequence 311, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 311:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-072-967-311

Query Match 6.9%; Score 55.6; DB 4; Length 1060;
Best Local Similarity 50.2%; Pred. No. 9.2e-07;
Matches 121; Conservative 22; Mismatches 95; Indels 3; Gaps 3;
Qy 338 ATGCGGCAAGCAATCTTGTCTATGCGGATGTTGTGATCGGTTTCGACGAAAGCCGGCATG 397
Db 744 ANNMGACCCNAAKATTTGNAMWGGGAAVNTMTGSGTTC-ACCAATTTCGGGGG 686
Qy 398 CCGGACCTTCGACCTCGACCGCCCTCTATTGAGACGAATTTGTGCGCACCAATCGATCT 457
Db 685 GGGNCCTTTGTMCTTGACCTTTTGTGTTTCNAGGSAATTS-KCCATNSNGGWWTT 627
Qy 458 TTACCGCGGTGAGCTTTTTCGACGCGCATCGGCCCTTACAACTCGCGCTACCGAGTCTGGG 517

Db 626 TACNCCGGGATTTTGGGCMATCTGTTCCCMCACTTCGCAACNGGTTCTKTCG 567
Qy 518 CGGACTGGGACTTCAATATTCGCTGCTCTCCAAACCGCGCTGATTACCCGCTACATGG 577
Db 566 CCGANTGGGANTTCAATATTCGCTGCTCTCCAAACCGCGCTGATTACCCGCTACATGN 508
Qy 578 A 578
Db 507 A 507

RESULT 8
US-09-072-596-325
; Sequence 325, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 325:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-596-325

Query Match 6.0%; Score 47.8; DB 3; Length 1022;
Best Local Similarity 49.1%; Pred. NO. 0.00022;
Matches 82; Conservative 20; Mismatches 63; Indels 2; Gaps 1;
Qy 1 ATGACTGGCCAGTGTCTTCGATAATTATCCCTACCTTCATGCGGTCGCGTACGCTGCAA 60
Db 405 WTGGCCGACCAWNNNTTYTCRATCWYCCCCACCTTAACTTGKTYGCSGTATTGCCT 464
Qy 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTTACCGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 465 KCCTGCCTCRACAGCMYCNCCCKTCAACCTCGCGTGACTCCAAGTGT--CTGGYCGAA 522
Qy 121 GCGGTTTCGACCGATCGGACCCCTGACATCGGAACAGTTCCTCGGCC 167

Db 523 SGGGGGYTCAMCGGACAAACCCCRANNTTCGCAAAATTTTCNCCCC 569
RESULT 9
US-09-072-967-330
; Sequence 330, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 330:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-072-967-330

Query Match 6.0%; Score 47.8; DB 4; Length 1022;
Best Local Similarity 49.1%; Pred. NO. 0.00022;
Matches 82; Conservative 20; Mismatches 63; Indels 2; Gaps 1;
Qy 1 ATGACTGGCCAGTGTCTTCGATAATTATCCCTACCTTCATGCGGTCGCGTACGCTGCAA 60
Db 405 WTGGCCGACCAWNNNTTYTCRATCWYCCCCACCTTAACTTGKTYGCSGTATTGCCT 464
Qy 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTTACCGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 465 KCCTGCCTCRACAGCMYCNCCCKTCAACCTCGCGTGACTCCAAGTGT--CTGGYCGAA 522
Qy 121 GCGGTTTCGACCGATCGGACCCCTGACATCGGAACAGTTCCTCGGCC 167
Db 523 SGGGGGYTCAMCGGACAAACCCCRANNTTCGCAAAATTTTCNCCCC 569

RESULT 10
US-09-902-540-4672
; Sequence 4672, Application US/09902540

Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 4672
LENGTH: 1332
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-4672

Query Match 5.5%; Score 43.8; DB 4; Length 1332;
Best Local Similarity 49.0%; Pred. No. 0.0042;
Matches 117; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 53 CGCTGCAAGCCTGCTCGGAAGCATCGTGGGCAGACCTACCGGGAAGTGGAAAGTGTC 112
DB 170 CCTGGAGAGTCTTCGGTGGCATGGGCGGAGCGGTGAGAGATCTCCGGCC 229
QY 113 TTGTCGACGGGGTTGACAGCATCGGACCTCGACATCGGAACAGTTTCGCGCCGGAAC 172
DB 230 TGGGAGGGGGCCCATCATCGACCCAGCGGCATCATGTCACCAATGACCACGTCATCC 289
QY 173 TCGGCTCGGACTGCTGCTTACAGCGGGCCGATGATGCGCCCTACGACGCGCATGAAC 232
DB 280 GGGGCGCTTCGCGCATCAAGTCATCTGGCGGATGGCGCTCGTTCGACGCGGAGGTCA 349
QY 233 GGGGCGCTGGCTAGCCAGCGCAATGGGTACTTTTTTTAGGCGCCGACGACACCTTC 291
DB 350 TCGGAGGACGCGGGAAGACCTAGCGGTGCTCAAGTCAATGCCAAGGAGGCCCTC 408

RESULT 11
US-09-902-540-1230
Sequence 1230, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1230
LENGTH: 24754
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-1230

Query Match 5.5%; Score 43.8; DB 4; Length 24754;
Best Local Similarity 49.0%; Pred. No. 0.016;
Matches 117; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 53 CGCTGCAAGCCTGCTCGGAAGCATCGTGGGCAGACCTACCGGGAAGTGGAAAGTGTC 112
DB 1780 CCTGGAGAGTCTTCGGTGGCATGGGCGGAGCGGTGAGAGATCTCCGGCC 1839
QY 113 TTGTCGACGGGGTTGACAGCATCGGACCTCGACATCGGAACAGTTTCGCGCCGGAAC 172
DB 1840 TGGGAGGGGGCCCATCATCGACCCAGCGGCATCATGTCACCAATGACCACGTCATCC 1899

QY 173 TCGGCTCGGACTGCTGTTCCACAGCGGCGCGATGATGGCCCTACGACGCGCATGAAC 232
DB 1900 GGGGCGCTTCGCGCATCAAGTATCTTGGCGAATGGCGCTCTGTTGACCGGAGGTCA 1959
QY 233 GCGGCGTGGCGTAGCCACAGCGGAATGGGTACTTTTTTTAGGCGCGGACGACACCTTC 291
DB 1960 TCGGAGGACGCGCGGAACGACCTAGCGGTGCTCAAGTCAATGCCAAGGAGGCCCTC 2018

RESULT 12
US-08-311-731A-132
Sequence 132, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 36412 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-132

Query Match 5.4%; Score 43.6; DB 4; Length 36412;
Best Local Similarity 50.5%; Pred. No. 0.022;
Matches 106; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 150 CGCGAAGAGTTTCGCGCGGAACCTCGGCTCGGACTGGTCTCACAGCGGCGCGATGA 209
DB 12241 CGCTATCTGTGCGCATGGTTGCAGACCTCGTGCCTGCTGAGCCTGCGCGCTGG 12300
QY 210 TGGGCGCTTACGACGCCATGAACCGCGCGTGGGTAGCCACAGCGGAATCGGTACTTT 269
DB 12301 TGATCGCTACCGTATGGGCACCCCGTTGCGCGACGCGGCGGATGAATGGAGCGTT 12360
QY 270 TTTAGGCGCCGACGACACCTCTTAGAACCAACACAGTTTGGCCCGAGGTAGCGCTTTCT 329
DB 12361 TCTCGCAAGGAGACCAACCTGTATGCTGTGAGACAGTCTGCTACCTAGTCTGCTTTGT 12420

Query Match 5.3%; Score 42.8; DB 1; Length 5859;
Best Local Similarity 48.7%; Pred. No. 0.017;
Matches 116; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 18:38:09 ; Search time 559 Seconds
(without alignments)
8482.492 Million cell updates/sec

Title: US-10-805-311-23
Perfect score: 801
Sequence: 1 atgactggccagtgtctc.....aacgaagcgcagaccctag 801

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	801	100.0	801	2 AAT74470	Aat74470 Open read
2	801	100.0	4435	2 AAT74479	Aat74479 Part of t
3	799.4	99.8	801	2 AAT74469	Aat74469 Open read
4	799.4	99.8	7995	2 AAT74478	Aat74478 Complete
5	346	43.2	768	8 ACA38491	Aca38491 Prokaryot
6	346	43.2	828	2 AAT74477	Aat74477 Open read
7	346	43.2	828	8 ACA40754	Aca40754 Prokaryot
8	346	43.2	32155	10 ADB74252	Adb74252 Mycobacte
9	346	43.2	110000	4 AAI9682_33	Continuation (34 o
10	346	43.2	110000	4 AAI9683_32	Continuation (33 o
11	346	43.2	110000	4 AAI9683_33	Continuation (34 o
12	249.8	31.2	110000	4 AAI9682_32	Continuation (33 o
13	60	7.5	900	8 ACA42153	Aca42153 Prokaryot
14	58.4	7.3	912	11 ABD13980	Abd13980 Pseudomon
15	57	7.1	34980	6 ABQ81845	Abq81845 Bifidobac
16	56.4	7.0	906	11 ABD17921	Abd17921 Pseudomon
17	55.6	6.9	1060	2 AAZ19211	Aaz19211 M. tuberc
18	55.6	6.9	1060	2 AAZ19423	Aaz19423 M. tuberc
19	55.4	6.9	1033	8 ACA48731	Aca48731 Prokaryot
20	53.8	6.7	1035	8 ACA51327	Aca51327 Prokaryot

21	50.8	6.3	2000	8 ADA71938	Ada71938 Rice gene
22	47.8	6.0	1022	2 AAZ19230	Aaz19230 M. tuberc
23	47.8	6.0	1022	2 AAZ19442	Aaz19442 M. tuberc
24	45.2	5.6	1827	11 ADM01454	Adm01454 Human cDN
25	45	5.6	37500	12 ADH48029	Adh48029 Clone FS3
26	43.6	5.4	36412	10 ADB74383	Adb74383 Mycobacte
27	42.8	5.3	4725	10 AAD55818	Aad55818 Micromono
28	42.8	5.3	5859	2 AAT14061	Aat14061 N. gonorr
29	42.8	5.3	5859	2 AAT49230	Aat49230 Lipo-Olig
30	42.8	5.3	60196	10 AAD55810	Aad55810 Micromono
31	42.6	5.3	1326	8 ACA35579	Aca35579 Prokaryot
32	42.6	5.3	1431	11 ACH95521	Ach95521 Klebsiell
33	42	5.2	110000	4 AAI9682_37	Continuation (38 o
34	41.4	5.2	1521	11 ABD13201	Abd13201 Pseudomon
35	41.4	5.2	2505	11 ABD12970	Abd12970 Pseudomon
36	41.2	5.1	1443	11 ABD14158	Abd14158 Pseudomon
37	40.8	5.1	320	3 ABQ62968	Abq62968 Mycobacte
38	40.8	5.1	882	8 ACA45716	Aca45716 Prokaryot
39	40.8	5.1	1836	3 AAC55818	Aac55818 S. lavend
40	40.8	5.1	1836	10 ADE10237	Adel0237 S. lavend
41	40.8	5.1	53500	3 AAC55842	Aac55842 Complete
42	40.8	5.1	53500	10 ADE10261	Adel0261 S. lavend
43	40.8	5.1	110000	4 AAI9682_01	Continuation (2 of
44	40.8	5.1	110000	4 AAI9683_01	Continuation (2 of
45	40.6	5.1	1185	8 ADA69780	Ada69780 Rice gene

ALIGNMENTS

RESULT 1
AAT74470
ID AAT74470 standard; DNA; 801 BP.
XX
AC AAT74470;
XX
DT 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 11-MAR-1998 (first entry)
XX
DE Open reading frame P from the GS region of M. paratuberculosis.
XX
KW GS; pathogenesis island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX
OS Mycobacterium avium subsp. paratuberculosis.
XX
FH Key Location/Qualifiers
CDS 1..801
FT /*tag= a

PN W09723624-A2.
XX
PD 03-JUL-1997.
XX
PF 23-DEC-1996; 96WO-GB003221.
XX
PR 21-DEC-1995; 95GB-00026178.
XX
PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX
PI Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
XX Ford J;
XX WPI; 1997-351061/32.
XX P-ESDB; AAW21775.
XX
PT New isolated pathogenicity island from mycobacteria - used to develop
XX products for detection, diagnosis, prevention and treatment of
XX mycobacteria infections.
XX
PS Claim 6; Page 50; 62pp; English.

XX The present sequence represents one of eight open reading frames (ORFs)
CC of a novel polynucleotide sequence designated "GS". GS is a pathogenicity
CC island of 8 kb of DNA comprising a core region of 5.75 kb with multiple
CC ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also
CC the transmissible element, encode proteins which may be linked to
CC pathogenicity, such as providing receptors for cellular recognition. GS
CC was discovered and characterised using differential DNA analysis
CC technology. It is found within Mycobacterium paratuberculosis and it has
CC also been identified in Mycobacterium avium subspecies silvaticum. These
CC pathogenic mycobacteria cause chronic inflammation of the intestine and
CC Crohn's disease in humans. The protein products of the ORFs of GS can be
CC used for detecting mycobacteria or for diagnosing, treating or preventing
CC mycobacterial diseases. In particular they can be used as vaccines for
CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
CC Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 801 BP; 168 A; 232 C; 228 G; 173 T; 0 U; 0 Other;

Query Match 100.0%; Score 801; DB 2; Length 801;
Best Local Similarity 100.0%; Pred. No. 6.2e-222;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTCGCCAGTCTTCGATATATTATCCCTACCTTCAATGCAGCGGTGACGCTGCAA 60
DB 1 ATGACTCGCCAGTCTTCGATATATTATCCCTACCTTCAATGCAGCGGTGACGCTGCAA 60
QY 61 GCCTCGCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGAAGTGGTCTTGTTCGAC 120
DB 61 GCCTCGCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGAAGTGGTCTTGTTCGAC 120
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGGAACAGTTTCGCGCCGGAACCTCGGCTCG 180
DB 121 GCGGTTTCGACCGATCGGACCTCGACATCGGGAACAGTTTCGCGCCGGAACCTCGGCTCG 180
QY 181 GCACGTGGTGTTCACAGCGGGCCGATGATGSCCTACGACGCGCATCAACCGGGCGTTC 240
DB 181 GCACGTGGTGTTCACAGCGGGCCGATGATGSCCTACGACGCGCATCAACCGGGCGTTC 240
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCCCTCTACGAACCA 300
DB 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCCCTCTACGAACCA 300
QY 301 ACCAGTTGGCCGAGTAGCGCTTTCTCGCGACCATGCGGGAAGCATCTTGTCTAT 360
DB 301 ACCAGTTGGCCGAGTAGCGCTTTCTCGCGACCATGCGGGAAGCATCTTGTCTAT 360
QY 361 GSCGATGTTGTATGCTGCTTCAGACGAAAGCGGCATGCGGACCTTTCGACCTCGACCGC 420
DB 361 GSCGATGTTGTATGCTGCTTCAGACGAAAGCGGCATGCGGACCTTTCGACCTCGACCGC 420
QY 421 CTCCTATTAGACGAAATTTGTGCCCAACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 421 CTCCTATTAGACGAAATTTGTGCCCAACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
QY 481 GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGCGGACCTGGGACTTCAATATTCG 540
DB 481 GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGCGGACCTGGGACTTCAATATTCG 540
QY 541 TGCTTCTCCAAACCGCGCTGATTAACCGCTACATGGACGCTGATTCGGAATACAAAC 600
DB 541 TGCTTCTCCAAACCGCGCTGATTAACCGCTACATGGACGCTGATTCGGAATACAAAC 600
QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGATAAAGATTTCAGAAAACGGCTGCCA 660
DB 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGATAAAGATTTCAGAAAACGGCTGCCA 660
QY 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTTCAGCGCGCATGCTGGCGTTTGTGAAGAC 720
DB 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTTCAGCGCGCATGCTGGCGTTTGTGAAGAC 720
QY 721 AAGGAGATCGCGCTTGGCTTGGTAGCGGTTGATAAGGGTTAAGGCGCTCTCCAAA 780

DB 721 AAGGAGATCGCGCTTGGCTTGGTAGCGGTTAAGGCGCTCTCCAAA 780
QY 781 GAACGAAGCGCAGAACCGTAG 801
DB 781 GAACGAAGCGCAGAACCGTAG 801
RESULT 2
ID AAT74479 standard; DNA; 4435 BP.
XX AAT74479;
XX 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 11-MAR-1998 (first entry)
XX Part of the GS region DNA sequence from M. paratuberculosis.
XX GS; pathogenicity island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX Mycobacterium avium subsp. paratuberculosis.
XX Key Location/Qualifiers
FH CDS 201..1232
FT /*tag= a
FT /note= "encodes AAW21769"
FT CDS 1172..2191
FT /*tag= b
FT /note= "encodes AAW21771"
FT CDS 2467..3189
FT /*tag= c
FT /note= "encodes AAW21773"
FT CDS 3335..4135
FT /*tag= d
FT /note= "encodes AAW21775"
XX WO9723624-A2.
XX 03-JUL-1997.
XX 23-DEC-1996; 96WO-GB003221.
XX 21-DEC-1995; 95GB-00026178.
XX (SGB0-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX Hermondaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
PI Ford J;
XX WPI; 1997-351061/32.
XX New isolated pathogenicity island from mycobacteria - used to develop
PT products for detection, diagnosis, prevention and treatment of
XX mycobacteria infections.
PS Claim 5; Page 43-44; 62pp; English.
XX The present sequence represents a novel polynucleotide sequence
CC designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising a
CC core region of 5.75 kb with multiple open reading frames (ORFs) and an
CC adjacent transmissible element of 2.5 kb. The ORFs, and also the
CC transmissible element, encode proteins which may be linked to
CC pathogenicity, such as providing receptors for cellular recognition. GS
CC was discovered and characterised using differential DNA analysis
CC technology. It is found within Mycobacterium paratuberculosis and it has
CC also been identified in Mycobacterium avium subspecies silvaticum. These
CC pathogenic mycobacteria cause chronic inflammation of the intestine and
CC Crohn's disease in humans. The protein products of the ORFs of GS can be
CC used for detecting mycobacteria or for diagnosing, treating or preventing

CC mycobacterial disease. In particular they can be used as vaccines for
CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
CC Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
CC (Updated on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 4435 BP; 904 A; 1289 C; 1288 G; 954 T; 0 U; 0 Other;

Query Match 100.0%; Score 801; DB 2; Length 4435;
Best Local Similarity 100.0%; Pred. No. 1.2e-221;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTTCGATATATATCCCTACCTTCAATGCGAGCGTGACGTGCAA 60

Db 3335 ATGACTGGCCAGTGTTCGATATATATCCCTACCTTCAATGCGAGCGTGACGTGCAA 3394

QY 61 GCTCTGCTCGAAGCATCTCGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTTCGAC 120

Db 3395 GCTCTGCTCGAAGCATCTCGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTTCGAC 3454

QY 121 GCGGTTTCGACCGATCGGACCTTCGACATCGGAACAGTTCCTCCGCCGGAACCTCGGCTCG 180

Db 3455 GCGGTTTCGACCGATCGGACCTTCGACATCGGAACAGTTCCTCCGCCGGAACCTCGGCTCG 3514

QY 181 CCACTGTGCTTCACAGCGGCGCCGATGATGSCCCCTACGACGCGCATGAACCGCGGCTC 240

Db 3515 CCACTGTGCTTCACAGCGGCGCCGATGATGSCCCCTACGACGCGCATGAACCGCGGCTC 3574

QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTTTATAGCGCGCGACACACCTCTACGAACA 300

Db 3575 GCGGTAGCCACAGGCGAATGGGTACTTTTTTATAGCGCGCGACACACCTCTACGAACA 3634

QY 301 ACCAGTTGGCCAGGTAGCCGCTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT 360

Db 3635 ACCAGTTGGCCAGGTAGCCGCTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT 3694

QY 361 GCGATGTTGATCGGTTGCGAAGAAAGCCGCGCATGCGGACCTTTGCACTCGAACGCG 420

Db 3695 GCGATGTTGATCGGTTGCGAAGAAAGCCGCGCATGCGGACCTTTGCACTCGAACGCG 420

QY 421 CTCCTATTGACAGCAATTTGCGACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480

Db 3755 CTCCTATTGACAGCAATTTGCGACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 3814

QY 481 GGCATCGGCGCTTACAACTCGCGTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG 540

Db 3815 GGCATCGGCGCTTACAACTCGCGTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG 3874

QY 541 TGCTTCTCAACCCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAA 600

Db 3875 TGCTTCTCAACCCGCGCTGATTAACCGCTACATGAGCGTGTGATTTCCGAATACAA 3934

QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGAGTGAAGAGTTCAGAAACCGGCTGCA 660

Db 3935 GACATGACCGGCTTCAGCATGAGGCGAGGAGTGAAGAGTTCAGAAACCGGCTGCA 3994

QY 661 ATGTACTTCTCGGTTTCGAGGTTGGAGACTTCGAGCGCATCTCGCGTTTGTGAAGAC 720

Db 3995 ATGTACTTCTCGGTTTCGAGGTTGGAGACTTCGAGCGCATCTCGCGTTTGTGAAGAC 4054

QY 721 AAGGAGATTCGCGCTTCGCGTTGCGTACGCGGTTGATAAGGGTTAAGCGGCTTCCTCAA 780

Db 4055 AAGGAGATTCGCGCTTCGCGTTGCGTACGCGGTTGATAAGGGTTAAGCGGCTTCCTCAA 4114

QY 781 GAACGAAGCGCAGAACCGGTAG 801

Db 4115 GAACGAAGCGCAGAACCGGTAG 4135

RESULT 3

AAT74469

ID AAT74469 standard; DNA; 801 BP.

XX AC AAT74469;

XX 17-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
XX 11-MAR-1998 (first entry)

DE Open reading frame F from GS region of *M. avium* subspecies *silvaticum*.

XX GS; pathogenicity island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.

XX Mycobacterium avium; subspecies *silvaticum*.

XX Key Location/Qualifiers
FT CDS 1..801
/*tag= a

PN W09723624-A2.

XX 03-JUL-1997.

XX 23-DEC-1996; 96WO-GB003221.

XX 21-DEC-1995; 95GB-00026178.

XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX Hermontaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
PI Ford J;

XX WPI; 1997-351061/32.

XX P-PSDB; AAW21774.

XX New isolated pathogenicity island from mycobacteria - used to develop
PT products for detection, diagnosis, prevention and treatment of
PT mycobacteria infections.

XX Claim 6; Page 50; 62pp; English.

XX The present sequence represents one of eight open reading frames (ORFs)
of a novel polynucleotide sequence designated "GS". GS is a pathogenicity
island of 8 kb of DNA comprising a core region of 5.75 kb with multiple
ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also
the transmissible element, encode proteins which may be linked to
pathogenicity, such as providing receptors for cellular recognition. GS
was discovered and characterised using differential DNA analysis
technology. It is found within *Mycobacterium paratuberculosis* and it has
also been identified in *Mycobacterium avium* subspecies *silvaticum*. These
pathogenic mycobacteria cause chronic inflammation of the intestine and
Crohn's disease in humans. The protein products of the ORFs of GS can be
used for detecting mycobacteria or for diagnosing, treating or preventing
mycobacterial disease. In particular they can be used as vaccines for
inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
(Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 801 BP; 167 A; 232 C; 229 G; 173 T; 0 U; 0 Other;

Query Match 99.8%; Score 799.4; DB 2; Length 801;

Best Local Similarity 99.9%; Pred. No. 1.8e-221;

Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTTCGATATATATCCCTACCTTCAATGCGAGCGTGACGTGCAA 60

Db 1 ATGACTGGCCAGTGTTCGATATATATCCCTACCTTCAATGCGAGCGTGACGTGCAA 60

QY 61 GCCTGCCCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTTCGAC 120

Db 61 GCCTGCCCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGAAGTGTCTTGTTCGAC 120

QY 121 GCGGTTTCGACCGATCGGACCTTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180

Db 121 GCGGTTTCGACCGATCGGACCTTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180

```
QY 181 CGACTGGTCTTCACAGCGGGCCCGATGATGGCCCCCTAGACGCCATGAACGGCGGCTC 240
Db 181 CGACTGGTCTTCACAGCGGGCCCGATGATGGCCCCCTAGACGCCATGAACGGCGGCTC 240
QY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACA 300
Db 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACA 300
QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360
QY 361 GCGGATGTTGATCGCTTCAGCAAAAGCCGCGATGCGGACCTTCGACCTCGACGCG 420
Db 361 GCGGATGTTGATCGCTTCAGCAAAAGCCGCGATGCGGACCTTCGACCTCGACGCG 420
QY 421 CTCCTATTGAGACCAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 421 CTCCTATTGAGACCAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
QY 481 GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC 540
Db 481 GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC 540
QY 541 TGCCTTCTCAACCCGCGCTGATTAACCGCTACATGAGCGTCTGATTTCCGAATACAA 600
Db 541 TGCCTTCTCAACCCGCGCTGATTAACCGCTACATGAGCGTCTGATTTCCGAATACAA 600
QY 601 GACATGACCGGCTTCAGATGAGGAGGAGGACTGATAAGAGTTTCAGAAAACGGCTGCCA 660
Db 601 GACATGACCGGCTTCAGATGAGGAGGAGGACTGATAAGAGTTTCAGAAAACGGCTGCCA 660
QY 661 ATGTACTTCTGGGTTGCGAGGTTGGAGACTTTCAGCGCATGCTGCGCTTTTGAAGAC 720
Db 661 ATGTACTTCTGGGTTGCGAGGTTGGAGACTTTCAGCGCATGCTGCGCTTTTGAAGAC 720
QY 721 AAGGAGATCGCGCTGCGCTTGGCTAGCGGTTGATAGGGTTAAGCGCGTCTCCAAA 780
Db 721 AAGGAGATCGCGCTGCGCTTGGCTAGCGGTTGATAGGGTTAAGCGCGTCTCCAAA 780
QY 781 GAACGAAGCGCAACCGTAG 801
Db 781 GAACGAAGCGCAACCGTAG 801

RESULT 4
AAT74478
ID AAT74478 standard; DNA; 7995 BP.
XX AC AAT74478;
XX DT 17-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 11-MAR-1998 (first entry)
XX DE Complete GS region DNA sequence from M. avium subspecies silvaticum.
XX GS; pathogenecity island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX OS Mycobacterium avium; subspecies silvaticum.
XX FH Key
XX FT Location/Qualifiers
XX FT 50. .427
XX FT /*tag= a
XX FT /note= "encodes AAW21766"
XX FT 772. .1605
XX FT /*tag= b
XX FT /note= "encodes AAW21767"
XX FT 1814. .2845
XX FT /*tag= c
```

```
FT CDS /note= "encodes AAW21768"
FT 2785. .3804
FT /*tag= d
FT /note= "encodes AAW21770"
FT 4080. .4802
FT /*tag= e
FT /note= "encodes AAW21772"
FT 4947. .5747
FT /*tag= f
FT /note= "encodes AAW21774"
FT 6176. .7042
FT /*tag= g
FT /note= "encodes AAW21776"
FT complement(6215. .7953)
FT /*tag= h
FT /note= "encodes AAW21777-78"
XX WO9723624-A2.
XX 03-JUL-1997.
XX 23-DEC-1996; 96WO-GB003221.
XX 21-DEC-1995; 95GB-00026178.
XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX Hermon Taylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
XX Ford J;
XX WPI; 1997-351061/32.
XX New isolated pathogenecity island from mycobacteria - used to develop
XX products for detection, diagnosis, prevention and treatment of
XX mycobacteria infections.
XX Claim 5; Page 40-43; 62pp; English.
XX The present sequence represents a novel polynucleotide sequence
XX designated "GS". GS is a pathogenecity island of 8 kb of DNA comprising a
XX core region of 5.75 kb with multiple open reading frames (ORFs) and an
XX adjacent transmissible element of 2.5 kb. The ORFs, and also the
XX transmissible element, encode proteins which may be linked to
XX pathogenecity, such as providing receptors for cellular recognition. GS
XX was discovered and characterised using differential DNA analysis
XX technology. It is found within Mycobacterium avium subspecies silvaticum
XX and it has also been identified in Mycobacterium paratuberculosis. These
XX pathogenic mycobacteria cause chronic inflammation of the intestine and
XX Crohn's disease in humans. The protein products of the ORFs of GS can be
XX used for detecting mycobacteria or for diagnosing, treating or preventing
XX mycobacterial disease. In particular they can be used as vaccines for
XX inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
XX Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
XX (Updated on 17-OCT-2003 to standardise OS field)
XX SQ Sequence 7995 BP; 1574 A; 2323 C; 2462 G; 1636 T; 0 U; 0 Other;
Query Match 99.8%; Score 799.4; DB 2; Length 7995;
Best Local Similarity 99.9%; Pred. No. 4.2e-221;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACTGCGCCAGTGTCTTCGATTAATTCCTACCTCAATGCGAGCGGTGACGCTGCAA 60
Db 4947 ATGACTGCGCCAGTGTCTTCGATTAATTCCTACCTCAATGCGAGCGGTGACGCTGCAA 5006
QY 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAGTCTCTTCGAC 120
Db 5007 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAGTCTCTTCGAC 5066
QY 121 GCGGTTTCGACCGCATCGGACCTCGACATCGGCAACAGTTTCGCGCCCGGAATCGGCTCG 180
Db 5067 GCGGTTTCGACCGCATCGGACCTCGACATCGGCAACAGTTTCGCGCCCGGAATCGGCTCG 5126
```

```
OY 181 GCACGTGCTGTTACAGCGGGCCCGATGATGCCCCCTACGACGCGCATGAACCGGGCGTC 240
Db 5127 GCACGTGCTGTTACAGCGGGCCCGATGATGCCCCCTACGACGCGCATGAACCGGGCGTC 5186
OY 241 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGGAGCAGACACCTCTACGAACCA 300
Db 5187 GCGGTAGCCACAGGCGAATGGGTACTTTTATAGCGCGGAGCAGACACCTCTACGAACCA 5246
OY 301 ACCAGTTGGCCCGAGGTAGCGCTTTTCTCGCGACCAATGCGGCAAGCATCTTGTCTAT 360
Db 5247 ACCAGTTGGCCCGAGGTAGCGCTTTTCTCGCGACCAATGCGGCAAGCATCTTGTCTAT 5306
OY 361 GCGGATGTTGATGCTGCTTCAACAGCAAAAGCCGCGATGCCGACCTTCGACCTCGACCGC 420
Db 5307 GCGGATGTTGATGCTGCTTCAACAGCAAAAGCCGCGATGCCGACCTTCGACCTCGACCGC 5366
OY 421 CTCCTATTGAGACGAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 5367 CTCCTATTGAGACGAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 5426
OY 481 GGCATCGGCCCTTACAACCTCGCTTACCGAGTCTGGGCGGACTTGGGACTTCAATATTGCG 540
Db 5427 GGCATCGGCCCTTACAACCTCGCTTACCGAGTCTGGGCGGACTTGGGACTTCAATATTGCG 5486
OY 541 TGCTTCTCAACCCGCGCTGATTACCCGCTACATGGACGTCGTGATTTCCGAATACAAC 600
Db 5487 TGCTTCTCAACCCGCGCTGATTACCCGCTACATGGACGTCGTGATTTCCGAATACAAC 5546
OY 601 GACATGACCGGCTTCAGCATGAGCGAGGGGACTGATAAGATTCAGAAAACGCTGCCA 660
Db 5547 GACATGACCGGCTTCAGCATGAGCGAGGGGACTGATAAGATTCAGAAAACGCTGCCA 5606
OY 661 ATGTACTTCTGGTTGACGGTGGGAGACTTGCAGGCGCATGCTGGCGTTTGTGAAGAC 720
Db 5607 ATGTACTTCTGGTTGACGGTGGGAGACTTGCAGGCGCATGCTGGCGTTTGTGAAGAC 5666
OY 721 AAGGAGATCCCGCTCGGCTTCGCTACCGGTTGATTAAGGGTTAAGGCGCTCTCCAAA 780
Db 5667 AAGGAGATCCCGCTCGGCTTCGCTACCGGTTGATTAAGGGTTAAGGCGCTCTCCAAA 5726
OY 781 GAACGAGCGGACGACCGTAG 801
Db 5727 GAACGAGCGGACGACCGTAG 5747
```

RESULT 5

ACA38491

ID ACA38491 standard; DNA; 768 BP.

XX AC ACA38491;

XX AC ACA38491;

XX 19-JUN-2003 (first entry)

XX DT

XX DE Prokaryotic essential gene #20148.

XX XX

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX XX Mycobacterium bovis.

XX OS

XX XX WO200277183-A2.

XX FN

XX XX 03-OCT-2002.

XX PD

XX XX 21-MAR-2002; 2002WO-US009107.

XX FF

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX XX

XX PA (ELIT-) ELITRA PHARM INC.

XX

PI

PI

XX

DR

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

SQ

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
P-PSDB; ABU34621.

New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 26361; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is one of the target
prokaryotic essential genes. Note: The sequence data for this patent did
not form part of the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

Sequence 768 BP; 142 A; 248 C; 217 G; 161 T; 0 U; 0 Other;

Query Match 43.2%; Score 346; DB 8; Length 768;
Best Local Similarity 71.0%; Pred. No. 6.8e-90;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

```
OY 1 ATGACTGGCCAGTGTCTCGATAATTAATCCCTACCTTCAATGACGGGTGACGCTGCAA 60
Db 1 ATGGCCGACCAATGTTTTCGATCATATCCCCACCTTGAACGGTGTGCGGTATTGGCT 60
OY 61 GCCTGCTCGGAAGCATGTCGGCGAGACCTTACCGGAAGTGAAGTGGTCTTGTGCGAC 120
Db 61 GCCTGCTCGACAGCATGCGCGTGCAGACCTTGCCTGACTTCGAGCTGGTACTGCTGAC 120
OY 121 GGGGTTTCGACCGGATCGGACCTTCGACATCGCAACAGTTCGCCCGGAACTCGGCTCG 180
Db 121 GGGGCTCGACGGACGAAACCTCGACATCGCAACATTTTCGCCCGCCCACTCGGCGAG 180
OY 181 CGACTGGTCTTCACAGGGGGCCGATGATGGCCCTTACGACGCGCATGACCGGGGCTC 240
Db 181 CGGTTGATCAATTCATCGCGACACCGCAGGGCGTCTTACGACGCGCATGAACCGGGGCTG 240
OY 241 GGGTAGCCACAGCGGAATGGGTACTTTTATAGCGCGGACGACACACCTCTACGAACCA 300
Db 241 GACCTGGCCACCGGAACGTGGTTCCTTTCTGGCGCGGACGACGCTGTACGAGGCT 300
OY 301 ACCAGTTGGCCCGAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCATCTTGTCTAT 360
Db 301 ACCAGTTGGCCCGAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCATCTTGTCTAT 360
```

Db 301 GACACCTGGCGGGTGGCGGCTTCATTGGCGAACACGAGCCGACGATCTGGTATAT 360
Qy 361 GCGATGTTGTCATCGTTCGACGAAAGCCGATGCGGACCTTCGACCTCGACCGC 420
Db 361 GCGGACGTGATCGATCGCTCAACCAATTTCCGCTGGGGTGGCGCTTCGACCTCGACCGT 420
Qy 421 CTCTTATTGAGACGAATTTGGCCACCAATCGATCTTTTACCAGCGGTGAGCTTTTCGAC 480
Db 421 CTGTTGTTCAAGCGCAACATCTGCATCAGGGATCTTCTACCGCGCGGACTCTTCGGC 480
Qy 481 GGCATGCGCCCTTACAACCTCGGTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
Db 481 ACCATCGGTCTCTACAACCTCGGTACCGGCTCTGGCGGACTGGGACTTCAATATTGCG 540
Qy 541 TGCCTTCTCAACCGCGCTGATTTACCGCTACATGAGCGTCTGATTTCCGATACGAC 600
Db 541 TGCCTTCTCAACCGCGCTGATTTACCGCTACATGAGCGTCTGATTTCCGATACGAC 600
Qy 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGATAAGAGTTTCAGAAAACGGCTGCGA 660
Db 601 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAAGCGGCTGCGG 657
Qy 661 ATGTAC 666
Db 658 ATGTCC 663

RESULT 6

AAAT74477
ID AAAT74477 standard; DNA; 828 BP.

AC AAAT74477;

XX 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX Open reading frame F from the GS region of M. tuberculosis.

XX GS; pathogenesis island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
XX

OS Mycobacterium tuberculosis.

FH Key Location/Qualifiers
FT CDS 1..828
FT /*tag= a

XX WO9723624-A2.

XX 03-JUL-1997.

XX 23-DEC-1996; 96WO-GB003221.

XX 21-DEC-1995; 95GB-00026178.

XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX HermonTaylor J, Doran T, Millar D, Tizard M, Loughlin M, Sumar N;
PI Ford J;

XX WPI; 1997-351061/32.

XX P-PSDB; AAW21783.

XX New isolated pathogenicity island from mycobacteria - used to develop
PT products for detection, diagnosis, prevention and treatment of
PT mycobacteria infections.

PS Claim 8; Page 56; 62pp; English.

XX The present sequence represents an open reading frame (ORF), ORF F, from
CC M. tuberculosis that has been found to have homology with ORF F of a novel
CC polynucleotide sequence designated "GS". GS is a pathogenesis island of

CC 8 kb of DNA comprising a core region of 5.75 kb with multiple ORFs and an
CC adjacent transmissible element of 2.5 kb. The ORFs, and also the
CC transmissible element, encode proteins which may be linked to
CC pathogenesis, such as providing receptors for cellular recognition. GS
CC was discovered and characterised using differential DNA analysis
CC technology. It is found within Mycobacterium paratuberculosis and it has
CC also been identified in Mycobacterium avium subspecies silvaticum. These
CC pathogenic mycobacteria cause chronic inflammation of the intestine and
CC Crohn's disease in humans. The protein products of the ORFs of GS can be
CC used for detecting mycobacteria or for diagnosing, treating or preventing
CC mycobacterial disease. In particular they can be used as vaccines for
CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
CC Johne's disease in animals. (Updated on 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 828 BP; 162 A; 264 C; 233 G; 169 T; 0 U; 0 Other;

Query Match 43.2%; Score 346; DB 2; Length 828;

Best Local Similarity 71.0%; Pred. No. 7e-90;

Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy 1 ATGACTGGCGGCGGCTTCGATTAATTCCTACCTTCATGCGCGGTGACGCTGCA 60
Db 58 ATGGCCGACCAATGTTTTCGATCATATCCCCACCTTGAACGTGGTATTCGCT 117
Qy 61 GCCTGCTCGGAAGCATCTCGGCGAGACCTACCGGAAGTGGAGTGGTCTTGTTCGAC 120
Db 118 GCCTGCTCGACAGATCGCCGCTCAGACCTCGCGTGAATTCGAGTGGTACTGGTCGAC 177
Qy 121 GCGGTTTCGACCGGATCGGACCTTCGACATCGCAACAGTTCCTCCCGGAACTCGGCTCG 180
Db 178 GCGGCTCGACGGACGAAACCTCGACATCGCAACATTTTCGCCCCCAACTCGGCGAG 237
Qy 181 CGACTGGTTCGACAGGCGGCGGATGATGGCCCCCTACGACGCGCATGAACGGCGGCTC 240
Db 238 CGGTTGATTCATTCATCGCACACCGGCGCTCTACGACGCGCATGAACCGGCGGTG 297
Qy 241 GCGCTAGCCACAGCGGAATGGGTACTTTTTAGCGCGCGACGACACCTCTACGAACCA 300
Db 298 GACCTGGCCACCGAAGCTGGTTCCTTCTGGCGGCGGACGACGCTGTACGAGCT 357
Qy 301 ACCAGTTTGGCCCGAGGTAGCGCTTTTCTGGCGAACCATGCGGCAAGCCATCTTGTCTAT 360
Db 358 GACACCTGGCGCGGTGGCGGCTTCATTGGCGAAACAGAGCCAGCGATCTGGTATAT 417
Qy 361 GCGCATGTTGATGCGTTCGACGAAAGCGGCGATCGCGACCTTCGACCTCGACCGC 420
Db 418 GCGCAGCTGATTCATGCGCTCAACCAATTTCCGCTGGGGTGGCGCTTCGACCTCGACCGT 477
Qy 421 CTCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 478 CTGTTGTTCAAGCGCAACATCTGCATCAGCGGATCTTCTACCGCGCGGACTCTTCGGC 537
Qy 481 GGCATCGGCGCTTACAACCTGCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
Db 538 ACCATCGGTCTCTACAACCTCGGTACCGGCTCTGGCGGACTGGGACTTCAATATTGCG 597
Qy 541 TGCCTTCTCAACCGCGGCTGATTTACCGCTACATGGACGCTCGTGAATTCGGAATACAC 600
Db 598 TGCCTTCTCAACCGCGGCTGATTTACCGCTACATGGACGCTCGTGAATTCGGAATACAC 657
Qy 601 GACATGACCGGCTTCAGCATGAGGCGGACTGATAAGAGTTTCAGAAAACGGCTGCGCA 660
Db 658 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAAGCGGCTGCGG 714
Qy 661 ATGTAC 666
Db 715 ATGTCC 720

RESULT 7

ACA40754

ID ACA40754 standard; DNA; 828 BP.

XX

RESULT 8	
ADB74252/c	
ID	ADB74252 standard; DNA; 32155 BP.
XX	
AC	ADB74252;
XX	
DT	04-DEC-2003 (first entry)
XX	
DE	Mycobacterium tuberculosis DNA.
XX	
KW	Non-naturally occurring peptide; anion pump protein; tuberculosis;
KW	hypersensitivity reaction; tuberculostatic; gene; ds.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	US6583266-B1.
XX	
PD	24-JUN-2003.
XX	
PF	16-SEP-1994; 94US-00311731.
XX	
PR	19-AUG-1993; 93US-00109181.
XX	

ADB74252/c
ID ADB74252 standard; DNA; 32155 BP.

AC ADB74252:

DT 04-DEC-20

Mycobacter

Non-native
WY
WV

yy
kw
hypergens

OS mycobacte

PN US6583266

PD 24-JUN-20

61-DES-91 Ad

PR 19-AUG-19

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *X. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences

SQ	Sequence	900 BP;	184 A;	325 C;	253 G;	138 T;	0 U;	0 Other;
Query Match	7.5%;	Score 60;	DB 8;	Length 900;				
Best Local Similarity	50.0%;	Pred. No. 7.3e-07;						
Matches	209;	Conservative	0;	Mismatches	200;	Indels	9;	Gaps
2;								
Qy	7	GCCCCAGTGTTCGGATATATTATCCCTACCTTCATTCGACCGGTGAGCTGCGAAGCTGC	66					
Db	46	GCCCCGCTGGTATCCGTCTGAGCGCCCTGCTTCACGCGGAAAAGTACTCTGGAGAAGCC	105					
Qy	67	CTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAAGTGTCTCTTGTGCGACGGCGGT	126					
Db	106	CTCGCGAGCATCTCAGACAGAGACTACCCGAAATTCGAAAGTGATCATCGTCGACAGCGGT	165					
Qy	127	TCGACCGATCGGACCTCTGCACATCGCGAACAGTTCCTCGCCCGGAACTCGGCTCGCGACTG	186					
Db	166	TCCACGCGAACACAGCTACCGCATGCTGGAGCAGTTGCGAAAGTCCACGGTTTTCCAGC--	223					
Qy	187	GTCGTTTCAGAGCGGGCCCGATGATGCCCCCTACGACGCCATGAACCGCGCGCTCGGGGTA	246					
Db	224	-----TCTACCGCAGCAGAACGAGGCGCTCAGCGGAGCGCTGAACTTCGGGCTTCGACAC	279					
Qy	247	GCACAGGGGAAATGGGTACTTTTTTTTAGCGCGCAGCAGACCCCTCTACGAAACCAACCG	306					

CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-creams, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (1) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
CC supplied by the European Patent Office
XX
SQ Sequence 349980 BP; 70780 A; 106600 C; 104724 G; 67876 T; 0 U; 0 Other;
Query Match 7.1%; Score 57; DB 6; Length 349980;
Best Local Similarity 52.7%; Pred. No. 4.8e-05;
Matches 148; Conservative 0; Mismatches 130; Indels 3; Gaps 1;
Qy 10 CAGGTGTTCTCGAATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAAGCCTGCCTC 69
Db 230637 CAGTGTGTCAGCATCATATTATCCCGGTGTATAAGGTGAGAGAGTTCTCTGACGAATGCGTG 230696
Qy 70 GDAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGACGGCGGTTCG 129
Db 230697 GGTCCGTTGTCTGCGCAGACGTACGCCAATCTGGAGATTCTGCTGGTGTGACGCGGCTCG 230756
Qy 130 ACGATCGGACCTCGACATCGGAAACAGTTTCCGCCCGGAATCTCGGCTCGCGACTGCTC 189
Db 230757 CCGACAAATTGCCCGGCCATGTGACGCTT---GGGCGCGCCGCGACCCGCGCATAGCG 230813
Qy 190 GTTCACAGCGGGCCGATGATGGCCCTTACGACGCCATGAACCGCGCGTCTCGCGGTAGCC 249
Db 230814 GTGATTCAAGCCCAACGCGGCGCTGTCCGACGCCCGCAACTCCGGTATCGCTGAGGCT 230873
Qy 250 ACAGCGGAATGGGTACTTTTATAGGCGCGCCGACACACCT 290
Db 230874 ATCGGCGGTACATATACTTCGCCGATTCCGACGACACCGT 230914

Search completed: March 6, 2005, 20:00:43
Job time : 566 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 18:39:13 ; Search time 3856 Seconds
(without alignments)
10065.516 Million cell updates/sec

Title: US-10-805-311-23
Perfect score: 801
Sequence: 1 atgactgcccagtggtctc.....aacgaagcgcagaaccctag 801

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	801	100.0	801	6	A63802	A63802 Sequence 23
2	801	100.0	4435	1	MAP223833	AJ223833 Mycobacte
3	801	100.0	4435	6	A63783	A63783 Sequence 4
4	801	100.0	301068	1	A8017231	AE017231 Mycobacte
5	799.4	99.8	801	6	A63800	A63800 Sequence 21
6	799.4	99.8	7995	6	A63782	A63782 Sequence 3
7	799.4	99.8	8938	1	MAS223832	AJ223832 Mycobacte
8	799.4	99.8	25861	1	AF125999	AF125999 Mycobacte
9	799.4	99.8	71286	1	AF143772	AF143772 Mycobacte
10	346	43.2	828	6	A63817	A63817 Sequence 38
11	346	43.2	31175	1	MTU00024	U000024 Mycobacteri
12	346	43.2	31176	1	MSG12	AD000009 Mycobacte
13	346	43.2	32155	6	AR345347	AR345347 Sequence
14	346	43.2	110000	1	AE000516	Continuation (33 o
15	346	43.2	110000	1	AE000516	Continuation (34 o
16	346	43.2	318050	1	BX248344	BX248344 Mycobacte
17	346	43.2	348676	1	BX842581	BX842581 Mycobacte
18	81.8	10.2	202301	1	AE017286	AE017286 Desulfovi
19	69.2	8.6	300425	1	AP005041	AP005041 Streptomy

20	67.8	8.5	302605	1	AE016938	AE016938 Bacteroid
21	67.6	8.4	110000	1	AP006841	Continuation (52 o
22	63	7.9	313800	1	SC093911	AL939114 Streptomy
23	62.2	7.8	110000	1	AP006840	Continuation (31 o
24	61.2	7.6	347750	1	AP002998	AP002998 Mesorhizo
25	60	7.5	900	12	AY658939	AY658939 Synthetic
26	60	7.5	1200	1	PAU70729	U70729 Pseudomonas
27	60	7.5	10007	1	AE004506	AE004506 Pseudomon
28	59.4	7.4	67651	1	AY523972	AY523973 Azospiril
29	59	7.4	738	9	HA334144	AJ334144 Homo sapi
30	59	7.4	20074	1	AB059427	AB059427 Gluconace
31	58.6	7.3	110000	1	RHE531985	Continuation (10 o
32	58	7.2	12498	1	AE004912	AE004912 Pseudomon
33	57.6	7.2	341887	1	AP003006	AP003006 Mesorhizo
34	57	7.1	10516	1	AE014688	AE014688 Bifidobac
35	57	7.1	349980	6	AX492783	AX492783 Sequence
36	57	7.1	349980	6	AX553950	AX553950 Sequence
37	55.6	6.9	1060	6	BD205988	BD205988 Compounds
38	55.6	6.9	1060	6	AR233268	AR233268 Sequence
39	55.6	6.9	1060	6	AR353473	AR353473 Sequence
40	54.8	6.8	303650	1	AP001519	AP001519 Bacillus
41	54.8	6.8	2582	1	AF292031	AF292031 Mesorhizo
42	54.2	6.8	345783	1	AP003001	AP003001 Mesorhizo
43	54	6.7	344805	1	EX640434	EX640434 Bordetell
44	54	6.7	348074	1	EX640449	EX640449 Bordetell
45	54	6.7	348134	1	EX640420	EX640420 Bordetell

ALIGNMENTS

RESULT 1
A63802
LOCUS A63802 801 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 23 from Patent WO9723624.
ACCESSION A63802
VERSION A63802.1 GI:3717374
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 801
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
1. 801
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03749.1"
/db_xref="GI:3717375"
/translation="MTAPVFSIIPTNAAVTLOACIGSVIGQYREVEVVLVDGGST
DRTLDIANSFRPELGSRLVHVSPPDGPYDAMNMGVATGEWVFLGADDTLYEPTT
LAQVAFLGHAASHLYGVGVVMSRKSAGPFDLDRLFLFETMLCHQSI FYRRLFD
GIGYNLRYRWADWDFNFCFSPNALITRYMDVVISEYNDMTGFSNRQGTDKFPRK
LPWFVWAGWETCERMLAFLKDKENRRLLALTRLIRYKAVSKERSABP"

ORIGIN
Query Match 100.0%; Score 801; DB 6; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.3e-160;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGACTGGCCAGTGTTCTCGATAATTATCCCTACCTCAATGACGGTGACCGTGCAA 60
|||||
Db 1 ATGACTGGCCAGTGTTCTCGATAATTATCCCTACCTCAATGACGGTGACCGTGCAA 60
|||||

QY	61	GCCTGCTCGGAGCATCGTGGGCGAGACCTACCGGGAAGTGGAGTGGTCTCTGTGCGAC	120
Db	61	GCCTGCTCGGAGCATCGTGGGCGAGACCTACCGGGAAGTGGAGTGGTCTCTGTGCGAC	120
QY	121	GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180
Db	121	GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG	180
QY	181	CGATGCTGTTTCACAGCGGGCCGATGATGGCCCTACGACGCGCATGAACCGGGCGTTC	240
Db	181	CGATGCTGTTTCACAGCGGGCCGATGATGGCCCTACGACGCGCATGAACCGGGCGTTC	240
QY	241	GCGTAGGACACAGGCGAATGGTACTTTTTTATGGCGCGACGACACCTCTACGAACCA	300
Db	241	GCGTAGGACACAGGCGAATGGTACTTTTTTATGGCGCGACGACACCTCTACGAACCA	300
QY	301	ACCACTTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	360
Db	301	ACCACTTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	360
QY	361	GCGCATGTTGTATCGGTTTCGACGAAAGCCGCGATGCCGACCTTTCGACCTCGACCGC	420
Db	361	GCGCATGTTGTATCGGTTTCGACGAAAGCCGCGATGCCGACCTTTCGACCTCGACCGC	420
QY	421	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC	480
Db	421	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC	480
QY	481	GGCATCGGCCCTTACAACTCGCGTACCGAGTCTGGCGGAGCTGGGACTTCAATATTCGC	540
Db	481	GGCATCGGCCCTTACAACTCGCGTACCGAGTCTGGCGGAGCTGGGACTTCAATATTCGC	540
QY	541	TGCTTCTTCCAAACCCGCGCTGATTAACCGCTACATGGAAGTTCGGAATACAAAC	600
Db	541	TGCTTCTTCCAAACCCGCGCTGATTAACCGCTACATGGAAGTTCGGAATACAAAC	600
QY	601	GACATGACCGGGTTGAGATGAGGCGAGGAGTGAATAAGAGTTTCAGAAAACGGTGC	660
Db	601	GACATGACCGGGTTGAGATGAGGCGAGGAGTGAATAAGAGTTTCAGAAAACGGTGC	660
QY	661	ATGTAATCTCTGGTTGCGAGGTTGGAGACTTTCAGCGCCATGCTGGCGCTTTTGAAGAC	720
Db	661	ATGTAATCTCTGGTTGCGAGGTTGGAGACTTTCAGCGCCATGCTGGCGCTTTTGAAGAC	720
QY	721	AAGGAGAAATCGCGTCTCGCTTCGATACCGGTTGATAAGGGTTAAGCGCTTCCAAA	780
Db	721	AAGGAGAAATCGCGTCTCGCTTCGATACCGGTTGATAAGGGTTAAGCGCTTCCAAA	780
QY	781	GAACGAGCGCAGAACCGTAG	801
Db	781	GAACGAGCGCAGAACCGTAG	801

RESULT 2
MAP223833 4435 bp DNA linear BCT 07-JAN-1999
LOCUS Mycobacterium avium paratuberculosis gs (ba,bb,c,d) genes.
DEFINITION
ACCESSION AJ223833
VERSION AJ223833.1 GI:3550477
KEYWORDS gsa gene; gsb gene; gsc gene; gsd gene.
SOURCE Mycobacterium avium subsp. paratuberculosis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE 1
AUTHORS Tizard,M., Bull,T., Millar,D., Doran,T., Martin,H., Sumar,N., Ford,J. and Hermon-Taylor,J.
TITLE A low G+C content genetic island in Mycobacterium avium subsp. paratuberculosis and M. avium subsp. silvaticum with homologous genes in Mycobacterium tuberculosis
JOURNAL Microbiology (Reading, Engl.) 144 (Pt 12), 3413-3423 (1998)

MEDLINE	99098703	
PUBMED	9884234	
REFERENCE	2 (bases 1 to 4435)	
AUTHORS	Bull,T.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-JAN-1998) Bull T., Surgery, St Georges Hospital Medical School, Cranmer Terrace, London, SW17 ORE, UK	
FEATURES	Location/Qualifiers	
source	1..4435	
	/organism="Mycobacterium avium subsp. paratuberculosis"	
	/mol_type="genomic DNA"	
	/sub_species="paratuberculosis"	
	/db_xref="taxon:1770"	
gene	201..1232	
	/gene="gsa"	
CDS	201..1232	
	/gene="gsa"	
	/function="unknown"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CA11575.1"	
	/db_xref="GI:3550478"	
	/db_xref="GOA:O86292"	
	/db_xref="Uniprot/TREMBL:O86292"	
	/translation="MKRALITGITQDGSYLAEILLSKGYEVHGLVRRASFTNTSRID HLYVDPQPGARLFHYADLTDCGTRLVTLSSIDPDEVYALAAQSHVRSFDEPVHTG DTTGMSIRLEAVRLSRVDCRFYQASSEMFGASPPQNESTFFYPRSPYGAAKVFS YWTRNRYEAYGLFVANGIILFNHESPRRGETFVTRKITRAVARIRAGVQSEVYWGND AIDWGYAPBYVEGMRMLQAPEDDYVLATGRGYTVREFAQAADFVGLDWQKHVF DRYLRPTVEDSLVGDADRAQAQSLGWKASVHTGELARIMVDADIAASECDGTPWIDTP MLPGWGSV"	
gene	1172..2191	
	/gene="gsb"	
CDS	1172..2191	
	/gene="gsb"	
	/function="unknown"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CA11576.1"	
	/db_xref="GI:3550479"	
	/db_xref="Uniprot/TREMBL:O86293"	
	/translation="MRWHTMDRHADVAWLGRSKLTTPGLDRATPVYIAGHRLGVGS ALVRRPRAEGFTNLIVRSRDEIDLTDRAAFTDFVSETRPOVIIDAAARVGGIMANNTY DAFVLSNLAIRIOTLLDAAVAVRPRLLFLGSSCIYKVAQPIHESALLTGPLEPTN DADAIAKIGLQVARRQIGLAWISAMPINLYGPDNFSFGSHLLPALIRYEEA KAGAEVTVNMGTPRRELLHVDLASACFLLEHFDGPNVNVNGTGVDSHISEIAD MVATAVGYIGETRWDPKDPGTPRKLLDVSALRELGLWRPRIALKDGIDATVSWYRTNA DAVRR"	
gene	2467..3189	
	/gene="gsc"	
CDS	2467..3189	
	/gene="gsc"	
	/function="unknown"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CA11577.1"	
	/db_xref="GI:3550480"	
	/db_xref="Uniprot/TREMBL:O86294"	
	/translation="MDFLRNAGLMARNVSTEMLRHFRKRLLVNQFKAYGVNVVIVDVG ANSGQFGSLRRAGFKSRIVSFEPISGPFQALTRFASDPLWCHQYALGADETITI NVAGNAGASSRVLPMKLSHODAPPANYIGTDVAIHRLDSVASEFLNPTDVTFLKID VQGFQKQVIAGSKSTLNESCUGMQLESLFPLYEGDMLIHEALSLVSLGFRLLGLLP GFTDFRNGRMQLQADGIFRGDD"	
gene	3335..4135	
	/gene="gsd"	
CDS	3335..4135	
	/gene="gsd"	
	/function="unknown"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CA11578.1"	
	/db_xref="GI:3550481"	
	/db_xref="GOA:O88109"	

/db_xref="UniProt/TREMBL:O88109"
/translation="MTAPVPSIIITFNAATVLRQACLSIVGQTYRVEVVLVDGGST
DRTLIANSFRELASRLVHSGDDPGYDAMNRGVATGEWVFLGADDTLYEPTT
LAQVAALFLVGVGDVVMRSTKGRHAGPFDRLPNTLNCHOSIFVRRLEFD
GIGPNLYRVWADWDENIRCFSPALITRYMDVVISYNDMTGFSMEQGTDKERKR
LPMYFWAGWETCRMLAFLLKDKENRLALRRLIRVXAVSKERSAEP"

ORIGIN		Query Match		100.0%; Score 801; DB 1; Length 4435;	
		Best Local Similarity		100.0%; Pred. No. 7.6e-160;	
		Matches 801; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGACTGCCCGCAGTGTCTCGATAATATATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA	60		
DB	3335	ATGACTGCCCGCAGTGTCTCGATAATATATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA	3394		
QY	61	GCCTGCTCGGAAGCATCGTCGGGCGAGCATCGCGGAAGTGGAGTGTCTTCTCGAC	120		
DB	3395	GCCTGCTCGGAAGCATCGTCGGGCGAGCATCGCGGAAGTGGAGTGTCTTCTCGAC	3454		
QY	121	GGCGGTTCGACCGATCGGACCTTCGACATCGCGAAGTGTCTCGCGCCGGAACCTCGGCTCG	180		
DB	3455	GGCGGTTCGACCGATCGGACCTTCGACATCGCGAAGTGTCTCGCGCCGGAACCTCGGCTCG	3514		
QY	181	CGACTGCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	240		
DB	3515	CGACTGCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	3574		
QY	241	GGCGTAGCCACAGGCGAATGGGTACTTTTATGAGCGCGCGAGCACACCTCTACGAACA	300		
DB	3575	GGCGTAGCCACAGGCGAATGGGTACTTTTATGAGCGCGCGAGCACACCTCTACGAACA	3634		
QY	301	ACCAGCTTGGCCGAGTACCGCTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	360		
DB	3635	ACCAGCTTGGCCGAGTACCGCTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	3694		
QY	361	GGCGATGTTGATCGGTTGCGAAGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	420		
DB	3695	GGCGATGTTGATCGGTTGCGAAGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	3754		
QY	421	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
DB	3755	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	3814		
QY	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTCGGCGGACTGGGACTTCAATATTCGC	540		
DB	3815	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTCGGCGGACTGGGACTTCAATATTCGC	3874		
QY	541	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGCGTCTGATTTCCGAATACAA	600		
DB	3875	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGCGTCTGATTTCCGAATACAA	3934		
QY	601	GACATGACCGGCTTACGATGAGGAGGAGGACTGATAAGAGTTTCAAAAACCGCTGCGA	660		
DB	3935	GACATGACCGGCTTACGATGAGGAGGAGGACTGATAAGAGTTTCAAAAACCGCTGCGA	3994		
QY	661	ATGTAATCTTGGGTTGAGGTTGGGAGTCTGACGCGCATGCTGGCGTTTGAAGAG	720		
DB	3995	ATGTAATCTTGGGTTGAGGTTGGGAGTCTGACGCGCATGCTGGCGTTTGAAGAG	4054		
QY	721	AAGGAGAAATCGCGCTTGGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCTCTCCAAA	780		
DB	4055	AAGGAGAAATCGCGCTTGGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCTCTCCAAA	4114		
QY	781	GAAACAGGCGAAGACCGTAG	801		
DB	4115	GAAACAGGCGAAGACCGTAG	4135		

RESULT 3
A63783 LOCUS
DEFINITION Sequence 4 from Patent WO9723624.
4435 bp DNA linear PAT 12-MAR-1998

ACCESSION A63783.1 GI:3717355
VERSION A63783.1
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 4 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
COMMENT Other publication AU 120797 19970717.
FEATURES
Location/Qualifiers
1..4435
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN		Query Match		100.0%; Score 801; DB 6; Length 4435;	
		Best Local Similarity		100.0%; Pred. No. 7.6e-160;	
		Matches 801; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGACTGCCCGCAGTGTCTCGATAATATATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA	60		
DB	3335	ATGACTGCCCGCAGTGTCTCGATAATATATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA	3394		
QY	61	GCCTGCTCGGAAGCATCGTCGGGCGAGCATCGCGGAAGTGGAGTGTCTTCTCGAC	120		
DB	3395	GCCTGCTCGGAAGCATCGTCGGGCGAGCATCGCGGAAGTGGAGTGTCTTCTCGAC	3454		
QY	121	GGCGGTTCGACCGATCGGACCTTCGACATCGCGAAGTGTCTCGCGCCGGAACCTCGGCTCG	180		
DB	3455	GGCGGTTCGACCGATCGGACCTTCGACATCGCGAAGTGTCTCGCGCCGGAACCTCGGCTCG	3514		
QY	181	CGACTGCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	240		
DB	3515	CGACTGCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGCTC	3574		
QY	241	GGCGTAGCCACAGGCGAATGGGTACTTTTATGAGCGCGCGAGCACACCTCTACGAACA	300		
DB	3575	GGCGTAGCCACAGGCGAATGGGTACTTTTATGAGCGCGCGAGCACACCTCTACGAACA	3634		
QY	301	ACCAGCTTGGCCGAGTACCGCTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	360		
DB	3635	ACCAGCTTGGCCGAGTACCGCTTCTCGCGGACCATGCGGCAAGCATCTCTGTCTAT	3694		
QY	361	GGCGATGTTGATCGGTTGCGAAGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	420		
DB	3695	GGCGATGTTGATCGGTTGCGAAGAAAGCCGCGATGCGGACCTTTTCGACCTCGACCGC	3754		
QY	421	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480		
DB	3755	CTCCTATTGAGACCAATTTGTCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	3814		
QY	481	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC	540		
DB	3815	GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC	3874		
QY	541	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGCGTCTGATTTCCGAATACAA	600		
DB	3875	TGCTTCTCCAAACCCGGCGTGAATACCGCTACATGAGCGTCTGATTTCCGAATACAA	3934		
QY	601	GACATGACCGGCTTACGATGAGGAGGAGGACTGATAAGAGTTTCAAAAACCGCTGCGA	660		
DB	3935	GACATGACCGGCTTACGATGAGGAGGAGGACTGATAAGAGTTTCAAAAACCGCTGCGA	3994		
QY	661	ATGTAATCTTGGGTTGAGGTTGGGAGTCTGACGCGCATGCTGGCGTTTGAAGAG	720		
DB	3995	ATGTAATCTTGGGTTGAGGTTGGGAGTCTGACGCGCATGCTGGCGTTTGAAGAG	4054		

Qy 721 AAGGAGAAATCGCGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
|||||
Db 4055 AAGGAGAAATCGCGCTTGGCTAGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 4114
|||||
Qy 781 GAACGAAGCGCAGAACCGTAG 801
|||||
Db 4115 GAACGAAGCGCAGAACCGTAG 4135

RESULT 4
AE017231
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..301068
/organism="Mycobacterium avium subsp. paratuberculosis
str. k10"
/mol_type="genomic DNA"
/strain="k10"
/sub_species="paratuberculosis"
/db_xref="taxon:262316"
complement(198..1433)
/locus_tag="MAP1150c"
complement(198..1433)
/locus_tag="MAP1150c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3467.1"
/db_xref="GI:41395600"
/translation="MALQOALLEVLDRNADADRDKQAETIYQALIDAEITAVI
GAGHERASRINQSRPKPTLSTIAGLELRIPKRSQFFPALLERRRRVDOCLF
AVVMEAYLHGTSTRKVDLVRKALGADAGISKEVSRI CADLDTVEGAFDRPLSEQHF
PYFDLDTYCKARVNHVRVSVQAVVIATGVAADGREVLFVGVGSEDCGAFWTAFLSL
KTRGLSGVLVSDAHTGLRGAIEAILGASWQRCRVHFLRNVLVAQVPKPSAEVMAAA
IRTIQAQPDABEVRQLODTIAGMLGRQLPKVETWLRREADDITAPDFPVLHKKIWS
TNPLRLNKEIKRRTDVGVPFNPANLRLAGSVLVAEHDEWQVAEKRYLSETTLLAL
HPRSDAQSVAVPAAITA"
complement(1792..2088)
/locus_tag="MAP1151c"
complement(1792..2088)
/locus_tag="MAP1151c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3468.1"
/db_xref="GI:41395601"
/translation="MSRPAPRSVPPAHHHEFAGRRSCPRVAETIARGYGLALRHDRP
ASPPAPGAAAGASRRFGEQVADIPLYLQIRHGEAVNGCRKAGVVPRAAIGE"
2109..3359
2109..3359
/locus_tag="MAP1152"
/locus_tag="MAP1152"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3469.1"
/db_xref="GI:41395602"
/translation="WDFGSLPEINSRIYSGPSAPLAAAAAHHGLAEHMSAAAS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

YGSAIELRTLWHGSPSTAMAAAAAFPIAWLGGTTAAQAEOTAAQATAAAAYDSVFAAT
VPPVIANRALLASLIATNVLGONTPALIATEAHYAEWMAQDAARMAYAYAGASAVAT
RLTFPGAPQSDANRAADQSAASALQLSTASSVESALSQGVSVQVPPVAQVATATV
TAAQLPLSLTDITGILKTFNSVMGTISGPTPLGVANLAKNWQIALUSIFSPVGTIQ
GIGLLHPKALTGLVAPLLRDLTGSTALSSAGTIVSASAGRALGLVLSVPAWNASA
VPVIRVAAELPETMLDAAPAMAVNGQGMFGPTALSSLAGRAVGGTATRAVASTVR
VPGAVAVDDLATTSTVIVIPNAK"
3371..4735
/locus_tag="MAP1153"
3371..4735
/locus_tag="MAP1153"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3470.1"
/db_xref="GI:41395603"
/translation="MVFSDFATLPPEVISTQIHVPGAVPILLAAAAWDGLAAELHGT
AASYASVISLWVSWQSSGMAAAAPYAWLSATAGLAEQTAQAQAAVAAAYEA
AAVATVPVVAANRSLATLQTNILGONDAATAATEDEYGOAWQADVAAMFSYAA
SESAGELTPEQAPATNAGPATQAAAAAESTSLSSWLKOLEAVITSLTGQYTOP
WQNLIGATGSTEIAPFEWETIYSSISQVGTQATVNVNATIGLVSGWKNFFIYAPW
SAAMKASLGLAGLASPGHIAAKPVSAGLVNAPTVGKLSVPAWATAAPALRLAS
TALPATSLAAAGATDFGLVNEATLGLAGGALGSPARVVTGTTGTPTRATGRRQG
PVKLDRIIAQLQEMPEQVQHNWVDEAGLDDLVARLRTTPGVHAVHVTDEQVAVGSPA
SKLG"
4792..5145
/locus_tag="MAP1154"
4792..5145
/locus_tag="MAP1154"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3471.1"
/db_xref="GI:41395604"
/translation="MKLLALLGVALGIVTAVPAQAVPEDEAATDNNNEVFIADLHK
VGISFQDPGQAVSAGKAVCVGLLARGSLQLLNDLRNPNALTITNGAQPATISAKSY
CPROLESAGCKGGG"
5238..6200
/locus_tag="MAP1155"
5238..6200
/locus_tag="MAP1155"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3472.1"
/db_xref="GI:41395605"
/translation="MDFGTAPESNGRHSHPGAGSLMDAAKAWGLGAQADALT
YVAVHIAANRLRRI SLAENCLGQLSPADADADAYERMWAQDAAMAYAVASAGV
WKLTPDSPPADPAPAGVTPSPERNWALTAPEVITTGROVISTISDALGALSVPL
ATFTLSLATGALSRLSSLSAPLDFALNHLNKAALGRAAVMCSGPRGGGGTN
SAALRVSGGAAIGILSVPRAWTRSPSPSPNSPAPLDRSQ"
6353..7747
/locus_tag="MAP1156"
6353..7747
/locus_tag="MAP1156"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO3473.1"
/db_xref="GI:41395606"
/translation="MKRLSSVDAAFNSAETAGWHVYCALAICDPKSAEYSFELRQ
LLIIRLPEVQOLRWVRWVAGLQDRPMFVEDEELDVDFHLRRIIGVPAPGGARRELELV
GLMSYKLDRLSRPLWEMVTEGLDGNAGGREFATLTKMHIAIVDVGSGAGLGEILLAT
PEPRPQOETVGSILVGFQLPGLERRAVGALINVGKTPFIARLEQTVRQOQVATLVG
RSRPPRYFEAPKTRFNAPVSPHRRVTCRVLARAKAIKDADVKLNDVRLVALVAGV
RDYLQKRGELPAKPLIAQIPVSTRTDENKDDIGNKVSMTVSLATHIEDPCRLEAIH
DSTQSAKLMAKALSAGQIMGLTETTPGLLHLAARAYTASGLSENLAIPINLVSNVPG
PPFLYIMAGAKLDSIGFPMDVALNITCFSTIDYDFGVFTTPEVANDIDMDRIEIP
ALTDLEKAAAGETG"
complement(7820..8524)
/locus_tag="MAP1157c"
complement(7820..8524)

QY	61	GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGAAAGTGGTCTCTGTCGAC	120
DB	91304	GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGAAGTGGAAAGTGGTCTCTGTCGAC	91363
QY	121	GGCGGTTGCAACCGATCGGACCTCGACATCGGGAACAGTTCCTCGCCCGGAACTCGGCTCG	180
DB	91364	GGCGGTTGCAACCGATCGGACCTCGACATCGGGAACAGTTCCTCGCCCGGAACTCGGCTCG	91423
QY	181	CGACTGGTGGTTCACAGGGGGCCGATGATGCGCCCTACGAGCGCATGAACCGCGCGTC	240
DB	91424	CGACTGGTGGTTCACAGGGGGCCGATGATGCGCCCTACGAGCGCATGAACCGCGCGTC	91483
QY	241	GGCGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAAACA	300
DB	91484	GGCGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAAACA	91543
QY	301	ACCAGTTGGCCAGGTAGCCGCTTTCTCGCGACCATGCGGCAAGCATCTTCTCTAT	360
DB	91544	ACCAGTTGGCCAGGTAGCCGCTTTCTCGCGACCATGCGGCAAGCATCTTCTCTAT	91603
QY	361	GGCGATGTTGTGATCGTTCCAGCAAAAGCCGACATGCGGACCTTTTCGACCTCGACCGC	420
DB	91604	GGCGATGTTGTGATCGTTCCAGCAAAAGCCGACATGCGGACCTTTTCGACCTCGACCGC	91663
QY	421	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	480
DB	91664	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC	91723
QY	481	GGCATGCGCCCTTACAACTCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC	540
DB	91724	GGCATGCGCCCTTACAACTCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC	91783
QY	541	TGCTTCTCCAAACCGCGGCTGATTTACCGCTACATGAGACGCTGATTTCCGAATACAAC	600
DB	91784	TGCTTCTCCAAACCGCGGCTGATTTACCGCTACATGAGACGCTGATTTCCGAATACAAC	91843
QY	601	GACATGACCGGCTTTCAGCATGAGCGGGAAGTCTGATAAGAGTTCAGAAAAACGGTGGCA	660
DB	91844	GACATGACCGGCTTTCAGCATGAGCGGGAAGTCTGATAAGAGTTCAGAAAAACGGTGGCA	91903
QY	661	ATGTAATCTCGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGG	720
DB	91904	ATGTAATCTCGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGG	91963
QY	721	AAGGAGATCGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGG	780
DB	91964	AAGGAGATCGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGG	92023
QY	781	GAACGAGCGCGAGAACCGTAG	801
DB	92024	GAACGAGCGCGAGAACCGTAG	92044
RESULT 5			
LOCUS	A63800	Sequence 21 from Patent WO9723624.	801 bp
DEFINITION	A63800		linear
ACCESSION	A63800		
VERSION	A63800.1	GI:3717372	
KEYWORDS		unidentified	
SOURCE		unidentified	
ORGANISM		unclassified.	
REFERENCE	1		
AUTHORS		Hermion-Taylor, J., Doran, T., Millar, D., Tizard, M., Loughlin, M., Sumar, N. and Ford, J.	
TITLE		NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY	
JOURNAL		Patent: WO 9723624-A 21 03-JUL-1997;	
COMMENT		ST GEORGE S HOSPITAL MEDICAL S (GB)	
FEATURES		Other publication AU 1202797 19970717.	
		Location/Qualifiers	

gene	locus_tag="MAP1157c"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AA03474.1"	
	/db_xref="GI:41395607"	
CDS	1	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
	91244	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
gene	locus_tag="MAP1158c"	
	/codon_start=1	
	/transl_table=11	
	/product="lipoprotein"	
	/protein_id="AA03475.1"	
	/db_xref="GI:41395608"	
CDS	1	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
	91244	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
gene	locus_tag="MAP1159c"	
	/codon_start=1	
	/transl_table=11	
	/product="lipoprotein"	
	/protein_id="AA03476.1"	
	/db_xref="GI:41395609"	
CDS	1	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
	91244	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
gene	locus_tag="MAP1160c"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AA03477.1"	
	/db_xref="GI:41395610"	
CDS	1	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA
	91244	ATGACTGCGCCAGTGTCTTCGATATATATCCCTACCTCAATGACGCGGTGACGCTGCAA

source	1. .801 /organism="unidentified" /mol_type="unassigned DNA" /db_xref="taxon:32644" 1. .801 /note="unnamed protein product" /codon_start=1 /protein_id="CAA03748.1" /db_xref="GI:3717373" /translation="MTAPVFSIIIPFNAVTLQCLSGIVGQTYREVVLVPGGST DRTLDIANSFRPELGRSLVHSGDDPGPYDMNRGVGATGEWVLFGLADDTLYEPTT LAQVAALFGRHAASHLVGDVVMRSTKSRHAGPFLDLRLLEFNLCHOSIFVRLELFD GIGPNLYRYWADNDENIRCFNSPALITRYMDVVISEYNDMTGFSMEQGTDKFERKR LPMYFWAGNETCRMJLAPLKDKNRRLALTRLRIRVAKVKSRAEP"	Db	781	GAACGAAGCGCAGAACCGTAG	801
CDS	RESULT 6 A63782 LOCUS A63782 7995 bp DNA linear PAT 12-MAR-1998 DEFINITION Sequence 3 from Patent WO9723624. ACCESSION A63782 VERSION A63782.1 GI:3717354 KEYWORDS SOURCE unidentified ORGANISM unidentified REFERENCE unclassified. 1 Hermion-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M., Sumar,N. and Ford,J. NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY Patent: WO 9723624-A 3 03-JUL-1997; ST GEORGE S HOSPITAL MEDICAL S (GB) Other publication AU 1202797 19970717. Location/Qualifiers 1. .7995 /organism="unidentified" /mol_type="unassigned DNA" /db_xref="taxon:32644"				
ORIGIN	Query Match 99.8%; Score 799.4; DB 6; Length 801; Best Local Similarity 99.9%; Pred. No. 1.6e-159; Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0; 1 ATGACTGGCCAGTGTTCGATAATATTCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60 Db 1 ATGACTGGCCAGTGTTCGATAATATTCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60 1 ATGACTGGCCAGTGTTCGATAATATTCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60 61 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTCTGTCGAC 120 Db 61 GCCTGCCCTCGGAAGCATCGTCGGGCAGACCTACCGGGAAGTGGAAGTGGTCTCTGTCGAC 120 121 GCGCGTTCGACCGATCGGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 180 Db 121 GCGCGTTCGACCGATCGGACCTCGACATCGGAACAGTTCCTCGCCCGGAACCTCGGCTCG 180 181 CGACTGGTGTTCACAGCGGGCCCGATGATGCGCCCTACGACGCGATGAACCGCGGCGTC 240 Db 181 CGACTGGTGTTCACAGCGGGCCCGATGATGCGCCCTACGACGCGATGAACCGCGGCGTC 240 241 GCGGTAGCCACAGCGCAATGGGTACTTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300 Db 241 GCGGTAGCCACAGCGCAATGGGTACTTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300 301 ACCAGTGTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360 Db 301 ACCAGTGTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT 360 361 GCGGATGTGTGATCGGTTCGACGAAAGCCGGGATGCGGACCTTTCGACCTCGACCGC 420 Db 361 GCGGATGTGTGATCGGTTCGACGAAAGCCGGGATGCGGACCTTTCGACCTCGACCGC 420 421 CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480 Db 421 CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480 481 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTGGGGGACTGGGACTTCAATATTCGC 540 Db 481 GGCATCGGCCCTTACAACCTCGGCTACCGAGTCTGGGGGACTGGGACTTCAATATTCGC 540 541 TGCCTTCCAAACCCGGCGCTGATTACCGGCTACATGAGCGTCTGATTTCCGAATACAA 600 Db 541 TGCCTTCCAAACCCGGCGCTGATTACCGGCTACATGAGCGTCTGATTTCCGAATACAA 600 601 GACATGACCGGCTTCAGCATGAGGAGCGGAGCTGATAAGAGTTCAGAAAACGGCTGCCA 660 Db 601 GACATGACCGGCTTCAGCATGAGGAGCGGAGCTGATAAGAGTTCAGAAAACGGCTGCCA 660 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTCGACGCGCATCTGCGCGTTTGTGAAGAC 720 Db 661 ATGTACTTCTGGGTTGCGAGGTGGAGACTTCGACGCGCATCTGCGCGTTTGTGAAGAC 720 721 AAGGAGATCGCGCTTGGCCCTTGGTACGCGGTTGATGAAGGGTTAAGCGGCTCCAAA 780 Db 721 AAGGAGATCGCGCTTGGCCCTTGGTACGCGGTTGATGAAGGGTTAAGCGGCTCCAAA 780 781 GAACGAAGCGCAGAACCGTAG 801				

/db_xref="GI:3550476" /db_xref="GOA:O88109" /db_xref="UniProt/TREMBL:O88109" /translation="MTAPVFSIIIPTFNAAVTLOACLGSIIVGQTYREVVVLDGGST DRLDIANSFRELGSRLVHSGDDGDPYDAMRGVGVATGEWLFILGADDTLYEPTT LAQVAAFGLDHAHLVYGDVVMRSTKSHAGPFDLRLLPETNLCHQSIIFYRLELFD GIGPNLYRYVMADWDNIRCFSPALITRYMDVVISYNDMTGFSMRQGTDEFRKR LPMYFWAGWETCRMLAPLXDKENRLALRLIRVKANSERSAEP"		/db_xref="GI:3550476" /db_xref="GOA:O88109" /db_xref="UniProt/TREMBL:O88109" /translation="MTAPVFSIIIPTFNAAVTLOACLGSIIVGQTYREVVVLDGGST DRLDIANSFRELGSRLVHSGDDGDPYDAMRGVGVATGEWLFILGADDTLYEPTT LAQVAAFGLDHAHLVYGDVVMRSTKSHAGPFDLRLLPETNLCHQSIIFYRLELFD GIGPNLYRYVMADWDNIRCFSPALITRYMDVVISYNDMTGFSMRQGTDEFRKR LPMYFWAGWETCRMLAPLXDKENRLALRLIRVKANSERSAEP"	
ORIGIN		ORIGIN	
Query Match 99.8%; Score 799.4; DB 1; Length 8938; Best Local Similarity 99.9%; Pred. No. 1.7e-159; Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		Query Match 99.8%; Score 799.4; DB 1; Length 8938; Best Local Similarity 99.9%; Pred. No. 1.7e-159; Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1	ATGACTGGCCAGTGTTCGATATATTCCTCAATGCAAGCGGTGACCGTGCAA	60
DB	5890	ATGACTGGCCAGTGTTCGATATATTCCTCAATGCAAGCGGTGACCGTGCAA	5949
QY	61	GCTCGCTCGGAAGCATGTCGGGCAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC	120
DB	5950	GCTCGCTCGGAAGCATGTCGGGCAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC	6009
QY	121	GCGCGTTCGACCGATCGGACCCCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG	180
DB	6010	GCGCGTTCGACCGATCGGACCCCTCGACATCGCAACAGTTCCTCGCCCGGAACCTCGGCTCG	6069
QY	181	CGACTGGTTCACAGCGGCGCCGATGATGCGCCCTACGAGCGCATGAACCGGGCGGTC	240
DB	6070	CGACTGGTTCACAGCGGCGCCGATGATGCGCCCTACGAGCGCATGAACCGGGCGGTC	6129
QY	241	GCGGTAGCCACAGGGCAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA	300
DB	6130	GCGGTAGCCACAGGGCAATGGGTACTTTTATAGCGCGCGACGACACCTCTACGAACCA	6189
QY	301	ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	360
DB	6190	ACCAGTTGGCCAGGTAGCGCTTTTCTCGCGACCATGCGGCAAGCATCTTGTCTAT	6249
QY	361	GCGCATGTTGTGATCGGTTGACGAAAGCCGGCATGCGGACCTTCGACCTCGACCGC	420
DB	6250	GCGCATGTTGTGATCGGTTGACGAAAGCCGGCATGCGGACCTTCGACCTCGACCGC	6309
QY	421	CTCCTATTGAGACCAATTTGTGCCACCAATCGATCTTTTACCGCGCGTGGCTTTTCGAC	480
DB	6310	CTCCTATTGAGACCAATTTGTGCCACCAATCGATCTTTTACCGCGCGTGGCTTTTCGAC	6369
QY	481	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG	540
DB	6370	GGCATCGGCCCTTACAACCTCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCG	6429
QY	541	TGCTTCTCCAACCCGGCGTGNATTACCGCTACATGGAAGTCTGATTTCCGAAATACAC	600
DB	6430	TGCTTCTCCAACCCGGCGTGNATTACCGCTACATGGAAGTCTGATTTCCGAAATACAC	6489
QY	601	GACATGACCGGCTTCAGCATGAGGACGAGGACTGATAAGAGTTCAGAAAAACGGCTGCCA	660
DB	6490	GACATGACCGGCTTCAGCATGAGGACGAGGACTGATAAGAGTTCAGAAAAACGGCTGCCA	6549
QY	661	ATGTACTTCTGGGTTGCAAGGTGGAGACTTTCAGGCGCATGCTGGCGTTTTTGAAGAC	720
DB	6550	ATGTACTTCTGGGTTGCAAGGTGGAGACTTTCAGGCGCATGCTGGCGTTTTTGAAGAC	6609
QY	721	AAGGAGATCGCGCTTCGCCCTTGGTAGCGGTTGATAAGGTTAAGCGCGTCTCCAAA	780
DB	6610	AAGGAGATCGCGCTTCGCCCTTGGTAGCGGTTGATAAGGTTAAGCGCGTCTCCAAA	6669
QY	781	GAACGAAGCGCAGAACCGGTAG 801	
DB	6670	GAACGAAGCGCAGAACCGGTAG 6690	
RESULT 8 AF125999		RESULT 8 AF125999	
LOCUS		LOCUS	
DEFINITION		DEFINITION	
ACCESSION		ACCESSION	
VERSION		VERSION	
KEYWORDS		KEYWORDS	
SOURCE		SOURCE	
ORGANISM		ORGANISM	
REFERENCE		REFERENCE	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
PUBMED		PUBMED	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	
TITLE		TITLE	
JOURNAL		JOURNAL	
AUTHORS		AUTHORS	


```

/gene="gtfa"
/!note="putative; similar to rhamnosyltransferase rtfa and
glycosyltransferase gtfb in this gene cluster"
/codon_start=1
/transl_table=11
/product="glycosyltransferase gtfa"
/protein_id="AAD20363.1"
/db_xref="GI:4416463"
/translation="MKCVLASVYGRDIEPSVVAARELQRRGHDMIMAVPPDVSFTE
AAGLAVPYGLDSQAWLDVYRNFTSFRFGWKGEMRMWMDLSDQWQWQNTT
LISLADGLDPAGOSYOEPAANVAEYDPLATLHVPMRPNQGVSVLPLAPLARS
MTAFWQWRLNKKVEDQRRRLGLPKATGSPRIARGLSLEIQAYDEVCPGLAE
WAWQDQRPFGVLTWELTDEADDEWMSWIRGTPPICFGSGMPVESPADTDMIGT
ACQIGERALLCAGSDSGVDPADPKVGVNYPATIPACRAVYVHGGSGTTAASL
RAGVPTLISMDANQITWGAQLKRLKVGATRFSAATDRDLSVQDLKRLAPDYARRR
ETAAARWPKADSVKMAADVVEDVQSRNTGRS"
4488..5774
/gene="rtfa"
/!note="rtfa"
/!note="similar to glycosyltransferases gtfa and gtfb in
this gene cluster"
/codon_start=1
/transl_table=11
/product="rhamnosyltransferase rtfa"
/protein_id="AAD20364.1"
/db_xref="GI:4416464"
/translation="MKFAVASVYGRDIEPCLAIGRELMMRGHSVRMAVPPNVDLAE
AAGLSAIVGDMDFWDETRDFPKNPLRGPTVLIAREWEPVLKNQWMSLMSV
GAGADLLTQYLQDLAINADYDGIPTWATLHYIPMRNGQLLPVPLVKTGNVY
DMWCMKKAEDRRRELGLPTATVAPRIARGLSLEIQAYHEVCPGLAEWAKA
GQRPFFVSLTMELEPTDDEVLISWIGQTPPICFGSGMPVAPESPAETIIEGTACA
ELGERALI.CSWSDFSNVPHADHVGVNYPATIPACRAVYVHGGSGTTAAGLRAG
IPSLTLTAGDQPLMGSRIKOLKVCTSRFSATTPETLAVDLKRLAPEYLTRARELS
SRMSKPASAGRAVDLLEFAPRSRGRCIPRAEQRSR"
5876..5876
/gene="mtfc"
/!note="mtfc"
/!note="putative; similar to methyltransferase mtfb in this
gene cluster and to Microspora griseorubida mycinamin
III-O-methyltransferase, PIR Accession Number S51595"
/codon_start=1
/transl_table=11
/product="methyltransferase mtfc"
/protein_id="AAD20365.1"
/db_xref="GI:4416465"
/translation="MTDHDTRFAYLDLRLRLTRYGSDLPVGLYRLGRPLFSTRNL
MLVRKRPKNKARDLGLDWPADALTMIGKRLTSLQHCVEVLEEDVPGDLVFCGVWR
GGASTLMRAVLAAYGDEKRCVWLCDSFAGVPPDPTVNYKADKGLRHRHARILGVLE
NVKANFERYGLLDQVRFPVPGWFKDLDKADPDRISVLRDLGLDYESTIQALDALYPR
LSPGGFCIVDDVYHAKACAQAVDVRTQGVTAIEVIDGTGLWRKP"
6776..7594
/gene="mtfb"
/!note="mtfb"
/!note="putative; similar to methyltransferases mtfb and
mtfc in this gene cluster"
/codon_start=1
/transl_table=11
/product="methyltransferase mtfb"
/protein_id="AAD20366.1"
/db_xref="GI:4416466"
/translation="MHADVMVAATVGLTPTGLRMWVRTEYWLARTLLPDVYANDA
LITFNHFAIDDEFPQRYRGRGSKEDWYQWPHVHVLGMLWAASASLSDGDFVE
CGSVGJUSSALDDELRIGKTYLDDTIFAGIDPFRVTVNRAAGALEISKGLRN
KANGVGVDFANFANVKNHRIIVGAVPETLDQVDAPSVAYLHIDMNCAPPEVAALRF
FWPRLTPGAFVLLDDYANRGRDEQRAMDEVAABFVGKICTLPTQGLLIKPAR"
7668..8447
/gene="dhga"
/!note="dhga"
/!note="putative; similar to the Streptomyces coelicolor

```

```

product encoded by the sequence presented in GenBank
Accession Number AL031350 and to Mycobacterium
tuberculosis Rv1544 and Rv2509"
/codon_start=1
/transl_table=11
/product="dehydrogenase dhga"
/protein_id="AAD20367.1"
/db_xref="GI:4416467"
/translation="MALAMITGASSIGLELAELFAORGYDLVAAENDGIYPASDSL
SRMGVDRPVQVDLRTPDGVEKLVREATEGGRGLDAALNAGVGGSGFVEGELADL
SIIDLAVSTHLAKLVLRNTRNTKVLTSLSLASMAGPPEAVYHASKSFVQSFSA
GHRKDELDTGITVLSMPGTETNFPFRAGNNQTRAGOKPDDPAEVARQGFEMMR
GHRKVGSLTGTGVVGRFLPDSWKAASRVNSTPAGRR"
complement(8454..9710)
/gene="gtfb"
complement(8454..9710)
/gene="gtfb"
/!note="putative; similar to glycosyltransferase gtfa and
rhamnosyltransferase rtfa in this gene cluster"
/codon_start=1
/transl_table=11
/product="glycosyltransferase gtfb"
/protein_id="AAD20368.1"
/db_xref="GI:4416468"
/translation="MRFALASYGTRGDIPEPSAAVGRELLRGHDVRLAVPPBLVGFVE
STGLAAVYPGKQVQSFDEEFLRNWRDFRNPIRLVKVMDPIIKYWADASATLKS
AEGADLLSTGLNFQAAANVAEYDIPLASLHHPMRANGOLVPMPSPLVRSWAAI
EMLFWRSTKVDNAQRQLGPKATRSORRIASHGLEIOAYDVCVPGLAENAHW
CDRPFVGCALTMELPTSDDDVASWIAAGTPPICFGSGSIPLESATVEMIGAACAR
LGERALICFGDIDVRGVDPHVKNVYSPVACRAVYVHGGSGTTAASLRAGI
PTLMASSADQPYWAAQIKRLKVGATRFSAATSETLVADLRLTILAPDYATRADLAA
RMTKPAESIETDADLLEARRKTLG"
9852..10505
/!note="low gc"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD20369.1"
/db_xref="GI:4416469"
/translation="MSRIGNSLVITIALNSARNKSLPTRLAAAYGWELGDHLVSPPA
IAKORHLNLFKSNRHRIFVEAGTYKQTTAFFIPHADQVSVLHDLGLFAAQRFA
QQPNVTLVRGSLVEIPRIVATSSPPLVDLGHFSGAGTAGEMEPASTLGLRLAD
VTPAGTTIVDDLRGLSGSLGSPFPQDITITGSARAAFPALIRTGLDSIVVEVPLGN"
complement(10567..11523)
/gene="hlpa"
complement(10567..11523)
/gene="hlpa"
/!note="putative; similar to Prevotella intermedia product
encoded by the sequence presented in GenBank Accession
Number AF052516"
/codon_start=1
/transl_table=11
/product="hemolytic protein hlpa"
/protein_id="AAD20370.1"

```

```

Query Match 99.8%; Score 799.4; DB 1; Length 25861;
Best Local Similarity 99.9%; Pred. No. 1.7e-159;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGACTGCGCCAGTGTTCGATAATTATCCCTTCAATGACGCGGTGACGCTGCAA 60
Db 17522 ATGACTGCGCCAGTGTTCGATAATTATCCCTTCAATGACGCGGTGACGCTGCAA 17581

Qy 61 GCCTGCTCGAAGCATGTCGGCGACGCTACCGGGAGTGGAGTGGTCTCTTCGAC 120
Db 17582 GCCTGCTCGAAGCATGTCGGCGACGCTACCGGGAGTGGAGTGGTCTCTTCGAC 17641

Qy 121 GGCGGTTTCGACCGCATCGGACCTCGACATCGCAACAGCTTTCCGCCCGGAACCTCGGCTCG 180
Db 17642 GGCGGTTTCGACCGCATCGGACCTCGACATCGCAACAGCTTTCCGCCCGGAACCTCGGCTCG 17701

Qy 181 CGACTGTTGTTTCACAGCGGCGCCGATGATGCGCCCTACGACGCGCATGAACCGCGCGTGC 240

```

Db	17702	CGACTGGTGTTCACAGCGGGCCGAGTGATGGCCCTACGACGCCATGAACGGCGGCTC	17761	TITLE JOURNAL	Direct Submission
Qy	241	GGCGTAGCACAGGCGAATGGGTACTTTTATAGGGCGGACGACACCTCTACGAACA	300	REFERENCE	Submitted (16-APR-1999) Microbiology, Colorado State University, Fort Collins, CO 80523-1677, USA
Db	17762	GGCGTGGCCACAGGGAATGGGTACTTTTATAGGGCGGACGACACCTCTACGAACA	17821	AUTHORS	4 (bases 1 to 71286) Eckstein,T.M., Lambert,M.L., Brennan,P.J., Belisle,J.T. and Inamine,J.M.
Qy	301	ACCAGTTGGCCAGGTAGCCCTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT	360	TITLE JOURNAL	Direct Submission
Db	17822	ACCAGTTGGCCAGGTAGCCCTTTCTCGCGACCATGCGGCAAGCCATCTTGTCTAT	17881	REMARK	Submitted (14-APR-2000) Microbiology, Colorado State University, Fort Collins, CO 80523-1677, USA
Qy	361	GGCGATGTGTGATCGGTTCGAGAAAGCGGCAATGCGGACCTTTTCACTCGACCGC	420	COMMENT	Sequence update by submitter
Db	17882	GGCGATGTGTGATCGGTTCGAGAAAGCGGCAATGCGGACCTTTTCACTCGACCGC	17941	FEATURES	On Apr 14, 2000 this sequence version replaced gi:5524305. Location/Qualifiers 1..71286 /organism="Mycobacterium avium" /mol_type="genomic DNA" /strain="2151" /db_xref="taxon:1764" 47..1132 /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAD44199.1" /db_xref="GI:5524306"
Qy	421	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTACCGCGGTGAGCTTTTCGAC	480	CDS	
Db	17942	CTCCTATTGAGACGAATTTGTGCCACCAATCGATCTTTACCGCGGTGAGCTTTTCGAC	18001		
Qy	481	GGCATCGGCCCTTACAACCTGGCTTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC	540		
Db	18002	GGCATCGGCCCTTACAACCTGGCTTACCGAGTCTGGCGGACTGGGACTTCAATATTCGC	18061		
Qy	541	TGCTTCTCAACCCCGCGCTGATTACCCGCTACATGCGCGTGTGATTTCCGAATACAAC	600		
Db	18062	TGCTTCTCAACCCCGCGCTGATTACCCGCTACATGCGCGTGTGATTTCCGAATACAAC	18121		
Qy	601	GACATGACGGGCTTCAAGATGAGGCGGAGTGAAGAGTTTCAAAAACGGCTGCGCA	660		
Db	18122	GACATGACGGGCTTCAAGATGAGGCGGAGTGAAGAGTTTCAAAAACGGCTGCGCA	18181		
Qy	661	ATGTACTTCTGGGTTGCAGGTTGGGAGACTTCAGCGCGATGCTGGCGTTTGTGAAGAC	720	repeat_region	1401..1408 /rpt_type=direct complement(1409..5670) insertion_seq="IS1601" complement(1524..2441) /codon_start=1 /transl_table=11 /product="IS1601-D" /protein_id="AAD44200.1" /db_xref="GI:5524307"
Db	18182	ATGTACTTCTGGGTTGCAGGTTGGGAGACTTCAGCGCGATGCTGGCGTTTGTGAAGAC	18241	repeat_region	
Qy	721	AAGGAGAAATCGCGCTCGCTTCGCTAGCGGTTGATAAGGGTTAAGCCGCTCTCCAAA	780	CDS	
Db	18242	AAGGAGAAATCGCGCTCGCTTCGCTAGCGGTTGATAAGGGTTAAGCCGCTCTCCAAA	18301		
Qy	781	GAAAGAGCGGACAGCCGTAG	801		
Db	18302	GAAAGAGCGGACAGCCGTAG	18322		
RESULT	9				
AF143772					
LOCUS					
DEFINITION					
AF143772					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					

repeat_region	/protein_id="AAD44203.1"	
	/db_xref="GI:5524310"	
CDS	/translation="MTAPHIVDPAGLLGEALAEASPDMLRSLQITVINALLSADADAV VGAEGQTPGRRAGNRGRHGLDTRVGTIDVAVPKLRKGTYPDMILRRKRAESA LITVADCYLAGVSTERNMDLKVTLTGIDLSKQSVNRATDLDEHVEAFRPLRDESSAG PFTFVAADLTKVREGGRVNVGVILLATGVNGDGRVGRVVTSTGPAWNEFFA DLVAGLGLRVLTSDAHAGLEAIAANLPGAARWCRTHYAANLMSICPKSMMPAVK AMLSVYQDPAPAAQDFRLIDYVTDKLPADVADLATAETIAFTAFPKVMSQI WSNPAERLKEIRRTDVGIFPNRDAIVRLVAGVLAEQNDWAEGRRLGLDLITR CRLTTIDNDPIDGANDPALTA"	
	5671..5678	
CDS	/rpt_type=direct	
	5791..6876	
gene	/note="similar to Bacteriophage MX1 A-protein encoded by GenBank Accession Number AF059242"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="A-protein-like protein"	
gene	/protein_id="AAD44204.1"	
	/db_xref="GI:5524311"	
CDS	/translation="MATELRKSLRRRTSERFGVAMRLARTIKPAETAAANSSGPLADLA GPNRPVALFVLGPORSCTSAITVLSICGCTLPALIGADANNPLGYWEPRAAISLN EAILRLGTNVDSPFRPLDQADESANTACINKIAIYLSLTPAALPLVYKEPRIT LSDLWFQARRAGDFVAARIVAPRQEVICIAKSWRVS PALASALMLKCNILAEKRT RGMKRVFYANLLDDMRRETKRISAEPLVELHTQDEDAIEKFLTPGLHRQHSQVST NRFGADWTSTVYALRDAAHDEPVDVDTATLDRVFDPSYRASERDPRKAHEDFHGLSNML FRVFRPAVAPVMEIFAMLRNRGTWA"	
	6996..8057	
CDS	/gene="gepia"	
	6996..8057	
gene	/note="TDP-D-glucose-dehydratase; putative; similar to Mycobacterium tuberculosis rmlB2 and galE"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="gepia"	
gene	/protein_id="AAD44205.1"	
	/db_xref="GI:5524312"	
CDS	/translation="MSASVLITGGAGFIGSLSRRLVEAGYDVALMDVLHPQVHGDR PVELAPSRLFTGVDPHADFDVAVLRFPRQIIVHLAETCTAQSLSRATPHQSVNVV GTTQLLDALSGSLVPEQLVLSRANVYEGAGWSGAEVFPQPRSHQLVAGRWDPQ GPAGQAVPLSPDRTEPRNTVYAAQLAQHLLAAWTAHDDLSVLRLQNVGP GQSLTNSYTGTVLAFARLAQOALEVEDGRIVRDFVYIDVVVDALFAVORPASQ RLRDVSGRATTIOELANTIAAMCEAPEPVVVGKFRDGDVRAASCIDIEPATSQLGWHF KWTLEDGLRALLEWIGNRp"	
	8321..9448	
CDS	/gene="mtfa"	
	8321..9448	
gene	/note="methyltransferase; putative; similar to mtfb and mtfC"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="MtfA"	
gene	/protein_id="AAD44206.1"	
	/db_xref="GI:5524313"	
CDS	/translation="MGIVDRTTVPVTKLRAPDRTRHMLWKAVLLGGALLAAACVF WPVLAVGIGAILLLLCARIPIGRDRDVIPLNYARDTPIYDDQYREFTRELAERLR RIGHTYLWEASQIPQPGAESEBLLDLGVMWIGNSTRLIFDTRHYIGDFTSGLV EDWLEDRIVKRGAPSLSEPPAFQPIRDTGTINDDGVPAALGRDVRPIKSGTYDTLA PFLADRPAAPIRLPHMDJDTYESCLHALETCKDHVVGSLVDFDEYLVINGEMAFYD FQKRYELSWQYVAGLEMIENMVTSRKRRLYSIAAIPGYLLGLDGRFLWACFRE PFRFWLNAADIEDPFIILGAAGSRKSVSIEITGLKGLAVPH"	
	9715..10536	
CDS	/gene="mtfb"	
	9715..10536	
gene	/note="methyltransferase; putative; similar to methyltransferase mtfC and Micromonospora griseorubida mycinamin III-O-methyltransferase encoded by GenBank Accession Number S51595"	
	/codon_start=1	
CDS	/transl_table=11	
gene	/product="MtfB"	
	/db_xref="GI:5524314"	
CDS	/translation="MTDRMRSPLDTRSAVLLDRLRLTGYSDDELVPVGMVYLGRP LFSTRKMLVRKPRFNHARDLGLDWPADALTMIGMORLTSLOHCVEITLKDIPGDL VECVGWSGASILMRAVLSAYDSKRCVMDLSDPEGVPPPTAHYQADQKILHRAAG ILAVPEAQVRANFERYLLDDRVFPGWFKDQAPRIERIALRLDGDYESTIOA LDALYPLRLSAGGICIIDYHAIDACROAVTDYRSEHGVTAPIEBIDGTGVLWRKS"	
	10658..11935	
CDS	/gene="gtfa"	
	10658..11935	
gene	/note="glycosyltransferase; putative; similar to rhamnosyltransferase rtfa and glycosyltransferase gtfb"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="Gtfa"	
gene	/protein_id="AAD44208.1"	
	/db_xref="GI:5524315"	
CDS	/translation="MKCVLASYTRGDIPEPSVVVARELQRRGHDMINAVPPDSVSFTE AAGTAVPYGLDSQALVDVYRNFWTSFPRGFMKMGMRMWRMMDLSDQCAQWNTT LLSLADGADILLFAGSQSYOEPAANVAEYDLPALTLHVPMPNGQVSVLPAPLARS MTAFDWCRLNKKVDAQRELGLPKATGSPRRIABERGSLEIQADVEVCFPLAAEE WAKWDGQRPFFVGTULMELITADDEVMWIRGGTPPICFGFSMPVSPADTVDMIGT ACAQLGSRALLCAGRSDFSGVPADPHKVVGVNYATIPACRAVWHHGGSGTAAASL RAGVPTLLISMDANQTTIWAQQLKELKLVGATRRFSATDRDLSLQDLRLKILAPDARRAR	
	99.8%; Score 799.4; DB 1; Length 71286;	
Query Match	Best Local Similarity 99.9%; Pred. No. 1.8e-159; Mismatches 0; Indels 0; Gaps 0;	
	Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 ATGACTGCGCCAGTGTCTTCGATAATTATCCTACCTTCAATGACGCGGTGACGCTGCAA 60	
	28153 ATGACTGCGCCAGTGTCTTCGATAATTATCCTACCTTCAATGACGCGGTGACGCTGCAA 28212	
Db		
Qy	61 GCCTGCTCGGAACATCGTCTGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC 120	
	28213 GCCTGCTCGGAACATCGTCTGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTTCGAC 28272	
Db		
Qy	121 GGCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCGGAACCTCGGCTCG 180	
	28273 GGCGGTTTCGACCGATCGGACCTCGACATCGCAACAGTTCCTCGCCGGAACCTCGGCTCG 28332	
Db		
Qy	181 CGACTGTGCTTTCACAGCGGGCCGATGATGGCCCTTACGACGCGCATGAACCGGGCGCTC 240	
	28333 CGACTGTGCTTTCACAGCGGGCCGATGATGGCCCTTACGACGCGCATGAACCGGGCGCTC 28392	
Db		
Qy	241 GGCGTAGCCACAGCGGAATGGGTACTTTTTCGCGCCGACGACACCTCTACGAACA 300	
	28393 GGCGTAGCCACAGCGGAATGGGTACTTTTTCGCGCCGACGACACCTCTACGAACA 28452	
Db		
Qy	301 ACCAGTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGAAGCCATCTTGTCTAT 360	
	28453 ACCAGTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGAAGCCATCTTGTCTAT 28512	
Db		
Qy	361 GGCGATGTTGTGATGCTGCTTCGACGAAAAGCGCGCATGCGGACCTTCGACCTCGACGCG 420	
	28513 GGCGATGTTGTGATGCTGCTTCGACGAAAAGCGCGCATGCGGACCTTCGACCTCGACGCG 28572	
Db		
Qy	421 CTCCTATTGAGACGAATTTGTGCGCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480	
	28573 CTCCTATTGAGACGAATTTGTGCGCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 28632	
Db		
Qy	481 GGCAATCGGCTTACACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTTCGC 540	
	28633 GGCAATCGGCTTACACCTCGCTACCGAGTCTGGCGGACTGGGACTTCAATATTTCGC 28692	
Db		
Qy	541 TGCTTCTCCAAACCGCGCTGATTACCGCTTACATGACGCTCGTGAATTCGGAATACAAAC 600	
	28693 TGCTTCTCCAAACCGCGCTGATTACCGCTTACATGACGCTCGTGAATTCGGAATACAAAC 28752	
Db		
Qy	601 GACATGACCGGCTTTCAGCATGACGAGGGGACTGATAAAGAGTTTCAGAAAACGGCTGCA 660	

```
Db 28753 GACATGACCGGCTTCAGCATGAGGAGGAGGACTGATAAGAGTTTCAGAAAACGGCTGCCA 28812
Qy 661 ATGTAATCTTGGTTGCGAGGTGGAGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAC 720
Db 28813 ATGTAATCTTGGTTGCGAGGTGGAGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAC 28872
Qy 721 AAGGAGAATCGCGCTCGGCTTGGTACGGGTTGATAAGGGTTAAAGGCGCTCTCCAAA 780
Db 28873 AAGGAGAATCGGCGTCTGGCTTGGCTACGGGTTGATAAGGGTTAAAGGCGCTCTCCAAA 28932
Qy 781 GAACGAAGCGAGAACCGTAG 801
Db 28933 GAACGAAGCGAGAACCGTAG 28953

RESULT 10
LOCUS A63817 828 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 38 from Patent WO9723624.
ACCESSION A63817
VERSION A63817.1 GI:3717387
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 38 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
COMMENT Other publication AU 120797 19970717.
FEATURES
source Location/Qualifiers
1..828
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
CDS
1..828
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03755.1"
/db_xref="GI:3717388"
/translation="MVQTKRYAGLTAAATKKVMAAPMFIIIPLTNVAAVLPACLDLS
IARQTCGFELVVDGGSTDELDIANIPENLGERLIIHRDQGVDAWNRGVDLRA
TGTWLLFLGADDSLYEADTLARVAAFIGHEPSDLVYGDVIMRSTNFRWGGAFDLRL
LFKRNICHOAIFYRGLFTIGPNLYRVLADWDFNIRCFSPALVTRYMHVVVASY
NEFGGLSNTIVDKELKPLKPLNSTRGLRLVILVLRWPKVISRAMVWRTVTSWRRR"

Query Match 43.2%; Score 346; DB 6; Length 828;
Beat Local Similarity 71.0%; Pred. No. 3.2e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy 1 ATGACTCGCGCAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
Db 58 ATGGCGCGACCAATGTTTCGATCATCATCCCCACCTTGACGTGGCTGCGGTATGGCT 117

Qy 61 GCCTGCTCGGAAGCATCGTCGGGCGAGACCTACCGGGAGAGTGGAAAGTGTCTCTGTCGAC 120
Db 118 GCCTGCTCGACAGCATCGCCGTCGAGACCTGCGGTGACTTCGAGCTGCTACTGCTGCGAC 177

Qy 121 GCGGTTTGAACGATCGGACCCCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 178 GCGGCTCGACGGAACCAACCTCGACATCGCAACATTTTCGCCCCCAACCTCGGCGAG 237

Qy 181 GCACGTGTGTTCACAGGGGGCCCGATGATGTCGCCCTTACGACGCGCATGAACCGGGCGGTC 240
Db 238 CGGTTGATCATTCATCGGGAACACCGAGCGGGCGTCTACGACGCGCATGAACCGGGCGGTG 297

Qy 241 GCGGTAGCCACAGGGGAATGGGTACTTTTTTTAGCGCGCGCAGCAGACCCCTCTACGAACCA 300
Db 298 GACCTGGCCACCGAACGTGGTTGCTCTTTCTTGGGCGCGGACGACAGCCTGTACGAGGCT 357
```

```
Qy 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCAACCATGCGCAAGCCATCTTGTCTTAT 360
Db 358 GACACCCCTGGCGGGTGGCGGCTTCATTGGCGAAACAGAGCCAGCGATCTGGTATAT 417
Qy 361 GCGGATGTTGTGATGTCGCTTCGACGAAAAGCCGGATGCGCGACCTTTTCGACCTCGACGCG 420
Db 418 GCGGAGTGTATCATGCGCTCAACCAATTCGCTGGGGTGGCGCTTCGACCTCGACCGT 477
Qy 421 CTCCTATTGTGAGCGAATTTGTGCGCACAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 478 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGCATCTTACCGCGCGGACTCTTCGCGC 537
Qy 481 GCGATCGGCTTACAACTCGGCTACGAGTCTGGGCGGACTGGGACTTCAATATTTCGC 540
Db 538 ACCATCGGTCCTCAACCTCCGCTACCGGGTCTCTGGCCGACTGGGACTTCAATATTTCGC 597
Qy 541 TGCCTTCTCCAAACCGGGCTGATTACCGCTACATGACGCTGCTGATTTCCGAAATACAAC 600
Db 598 TGCCTTCTCCAAACCGAGCGCTGTCACCGCTACATGACGCTGCTGCTTGCACGCTACAAC 657
Qy 601 GACATGACCGCTTCAGCATGAGGAGGGAAGTCTATAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 658 GAATTGCGCGGCTCAGCAATACGATCG---TCGACAAAGAGTTTTCAGACGGGTGCGC 714
Qy 661 ATGTAC 666
Db 715 ATGTCC 720

RESULT 11
LOCUS MTU00024/c 31175 bp DNA linear BCT 11-MAR-2002
DEFINITION Mycobacterium tuberculosis cosmid tbc2.
ACCESSION U00024
VERSION U00024.1 GI:560506
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
AUTHORS Smith,D.R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 31175)
AUTHORS Robison,K.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1994) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
COMMENT On Oct 27, 1994 this sequence version replaced gi:414230.
This sequence data was produced by the Genome Sequencing Center
located at Genome Therapeutics Inc. (formerly Collaborative
Research Inc.) (100 Beaver St., Waltham MA, 02154 617-893-5007).
Please contact Doug Smith (smith@cric.com) for further
information. The annotation should be considered preliminary and
incomplete.
FEATURES
source Location/Qualifiers
1..31175
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
<1..902
/codon_start=3
/transl_table=11
/product="u0002a"
/protein_id="AAA50950.1"
/db_xref="GI:560529"
/translation="SPRCARTLAGFESRLACRRYARGAVALTAATSRQIEHVRISET
VADSKPPPRDLKSDGSFGLVQIGTTIVVVFVAVLVFYIIVSRDDKDGVAAPG
DAVRTSKLVTPQTSNPKAVSFYEDFLCPAGCIFERGSGPTVSKLVDIGAUAADY
TMAILDSASNQHSYSSRAAAAYCADESIEAFRRFHAALFSKDIQPAELGKDPDA
RLIELAREAGVVGKVPKVCINSKGYIEKVDGLAAAVNVHATPTVRVNGTIEYEWSTPAAL
```

CDS
VAKIKEIVCDVDPGIDSAATATS"
908..1540
/codon_start=1
/transl_table=11
/product="u0002b"
/protein_id="AAAS0949.1"
/db_xref="GI:560528"
/translation="MVAARPAERSGDPAARVPVPSAWVVLIGGVIGLFPASMTLTVK
VRLDPTVPSNVPNTVSCSVNTTQASLLGPNPLGAGTFTVVVTVGLAVAK
VPLRWYVIGLAVGVAFVHMLIFQSLRIGALCPYCMVMAVIAATLVASIV
FGPRENGSERVCARLLYQWRWSLATLMTFTTFLIMVRFWDYNSTLI"
1565..4939
/codon_start=1
/transl_table=11
/product="pvc"
/protein_id="AAAS0948.1"
/db_xref="GI:560527"
/translation="MFSKVLVANRGEIAIRAFRAAYELGVGTAVVYVEDRNSQHLK
ADESQIGDIGHVPHAYLSVDEIVATARRAGADAIYFGYFELSENPDIAAACAAGIS
FVGPSEVLELAGNKSRAIAAREAGLFLMSAFVSDELLSVAAAGMPFLFYKAV
AGGGGRMVRVGDIAALPEAIEAASREAESAFGDPVYLEQAVINPHIEVQILADNL
GDVHLXERDCSVORRHOKVIELAPAPHLDAELRYKCMVDAAVAFARHIGYSCAGTVEF
LIDEGEVYFLEMPPRQVETHTVEITDVLVASQLRIAAGETLEQLGLQEDIAFH
GAAQCRIITDPANGFRPRAGSARCDPPAPVPSAWTAAPTWRKNQFVRLRHAGQAD
LSGRDLTPASRRARIAEFRIKGVSTNIPFLQAVLDDPDPFRAGVTTSPIDERPQLL
TARASDRGTIKLNPADVTNNPYGSRPTIYDDKLDPDLRAAPPAGSKQRLVKL
GPEGFAWLRESAAVGVDTTFRDAHQSLLATRVTSGLSVAPYLAKMTQQLLSVEC
WGAATYDVALRFLKEDPWERLATRAAMNCLQMLLRGNTVGTVPYPIVTSAFVQ
EATATGIDIFRIFDALNIESMRPAIDAVRETSGEIAEVAMCYGDLTDPEGEQIYTL
YTLKLEQIVDAGHVLAIKDMAGLLRPPAQRVLSALSRPDLPHVHTDHTFGQL
ASYVAWHAGADAVGAAPLAGTTSQALSIVAAAHTEYDGLSLVSALEAPYV
EALRYVAFESGLPGPTGVYHHEIPGQSLNFKQAIALGLGDRFEEIEEAVAGAD
RVLGRLVPTKVVGDALALVAGAVSDEFADPARFGIPESVLEFLREGLGDDP
GWPPEPLTALAGARPTAQLAADDEIALSSVGAKRQATLNEILLPPSPKSENEH
REAVGDTSQLANQFPYGLRQGEHVRKLERGVELLIGLEAISPEDEGRMTVMCILN
QOLRPVLRDTSIASAVPAEAKRGNFCHTAAAPPAGVTVGVGVGVAGVAGQTIATI
EAMKMEAPITAPAVAGTVERVAVSDTAEWRAETCWWV"
5057..5704
/codon_start=1
/transl_table=11
/product="u0002d"
/protein_id="AAAS0947.1"
/db_xref="GI:560526"
/translation="MQGYRPLAGDEAMRWGRYPLARESGLDPPHRRRRRIAYP
PRGTRPTDVRNGLSFLVARRDLTGLAVLDLVAGSGALGLEALSRGAASVLFVESD
QRSAVIARNTREALGLSATLRGCAVAVAAGTTSVDVLADPPVNVDSADVDAIL
AALGTNGTRGTAVVERATCAPLTWPEGRRWRPQRYVYGDTRLELAERLFANV"
5790..6275
/codon_start=1
/transl_table=11
/product="u0002e"
/protein_id="AAAS0946.1"
/db_xref="GI:560525"
/translation="MTGAVCPGSGFDPVTGLHVDIIFERAAAQFDEVVAILVNPAGTGM
FDLHERIAVMSVETPLNLRVQVHGLVDFVRSCTGTAIVKGLTGCTDPEYELQWA
QMNKHIAGVDTFFVATADRYSFVSSSLAKEVAMLGDDVSELLPEPVNRRRLDRLNT
T"
complement(7144..8076)
/codon_start=1
/transl_table=11
/product="purU"
/protein_id="AAAS0945.1"
/db_xref="GI:560524"
/translation="MKGSMTHAHATPNEPDYPPPPGPPPPADIGRLLLRCHDRPGII
AAVSTFLARAGANIISLDQHSAPEGGTFLOAIFHLPGLTAAVDELQDRDPSVADK
FGIDYRFAEAKRVAIMASTEDHCLDLRLNRNREGLEMSVVMVIANHPDLAAHVR
PFGVPFTHIPATRDTRAEQRLQSLGNDVLVLYARMQILSPGLEAIGCPLINI
HHSFLPAPTGAAPYORAREGVKLIATATVYVTEVLDBEGPIIEQDVVRVDHTHTVDLL
VRVGVADVERAVLSRAVLWHCODRVIVHNNQIVF"
complement(8149..9369)
/codon_start=1
/transl_table=11
/product="u0002f"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translation="MTSTKVEDRVTAAVLGAIGHALALTASMTWEILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATIATGLGAASSCSVAVALASLFRKGANFTAAAM
AFEGISTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAVFLFLFGVARGALIDARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLISPPGFTSIHAHVFVMEWLAILRDLILGL
LIAGALAAWPESPFQWFFLANHPAWSWVGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GGIFGGVIAFIFADLLIPLINITYKYTGAMMLVLGLTGYASVMVAGYITELLEFGT
TNLISQORSATVMTABEISWNTTWNVIFLVAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDHDDHDLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translation="MRASRLPALLPSLGPPLWLSRSALHHIPGGIYTVGVGNDRVAV
PLNVSPHSCGFPFRVSVCIYATASRGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAEAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHAIHTIPS
ERFENGLTQGRFYAMRTLKVEADRLVDEIAPDLVVGDLRISLSVSRHLAGIPIA
IANAWSPGAORRPLPDVITWTRLFGVRLVKLLYRLERPLLPALQCMPLNWWRRHGL
SSLCWNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVYATLGTSGGRNLLQLVLNALAELFVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translation="MPANAFVADYLPGEAAAARSADVVCNGGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGVMQVISRSEYPTSRERATRCLR
TRPWIIPAAECGCEPHAGEPAPDMVGQLVTST"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translation="MGLVRSRTSLVGLIGLVLVASFPAQLFYRPSDAVAEEVHKW
YVGNLVTKTYMGINCWKSVDMMNYQELISELQPSLVIEFGTRYGGSVAVYANIMR
QIGQFPKVLVDNSHKALDPPARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVEDSSINGHNGHVPFGFGPGPYEAIEYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translation="MEETSVAGDPDPAGTSTAPNAPEPVARRQRIILFVGEAATLH
VREPVLSRLDPSREYVHACDPRENKLLGPLPFPHPHITHTVPSEVLLKTAQGRLF
YNTTLRKTYAADRKIINEIADPVVVGDNRLSLVSARLACIPIAIAANATWSPARR
RFPDPDPTFRFFGVRPVSILYRLRYLFALYCLPLNWLRRKHGSLSLGWDLCRIPT
DGDYTLADVPELVTYTNLPANHYRLGPVLMSPDKPPTWWHSLPTDRPIIYATIGSS
GGKTLLOVNLALADLPVITVIAAGRNHLKNVPANAFVADLPGEAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSNMDQHLNMEALERAGAVLLRTERLNTGEGVAAAQVQ
LSGAEFRQARRLPKPSDQTLPSRSTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translation="MVQTKRYAGTAAANTKKVAAWAPMFSIIIFTLNVAAYLPACLDS
IARCTGDFELVLDGGSTDETDIANIFAPNLGERLIIRHDTQGVVDANRWGVDLA
TGLVFLGADDSLYEARVAFTGEHPSDLVYGDVIMRSTNFRWGCAGFOLDRL
LFKENIQAIFVYERGLFGTIGPVNARVRLADWDNENRCSFNALVRYHVVVAVSY
NEFGGLSNTIVDEKFLKRLPMSTLGRILVILVVRPKVISRAMWRTVLSWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translation="MTSTKVEDRVTAAVLGAIGHALALTASMTWEILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATIATGLGAASSCSVAVALASLFRKGANFTAAAM
AFEGISTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAVFLFLFGVARGALIDARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLISPPGFTSIHAHVFVMEWLAILRDLILGL
LIAGALAAWPESPFQWFFLANHPAWSWVGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GGIFGGVIAFIFADLLIPLINITYKYTGAMMLVLGLTGYASVMVAGYITELLEFGT
TNLISQORSATVMTABEISWNTTWNVIFLVAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDHDDHDLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translation="MRASRLPALLPSLGPPLWLSRSALHHIPGGIYTVGVGNDRVAV
PLNVSPHSCGFPFRVSVCIYATASRGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAEAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHAIHTIPS
ERFENGLTQGRFYAMRTLKVEADRLVDEIAPDLVVGDLRISLSVSRHLAGIPIA
IANAWSPGAORRPLPDVITWTRLFGVRLVKLLYRLERPLLPALQCMPLNWWRRHGL
SSLCWNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVYATLGTSGGRNLLQLVLNALAELFVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translation="MPANAFVADYLPGEAAAARSADVVCNGGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGVMQVISRSEYPTSRERATRCLR
TRPWIIPAAECGCEPHAGEPAPDMVGQLVTST"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translation="MGLVRSRTSLVGLIGLVLVASFPAQLFYRPSDAVAEEVHKW
YVGNLVTKTYMGINCWKSVDMMNYQELISELQPSLVIEFGTRYGGSVAVYANIMR
QIGQFPKVLVDNSHKALDPPARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVEDSSINGHNGHVPFGFGPGPYEAIEYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translation="MEETSVAGDPDPAGTSTAPNAPEPVARRQRIILFVGEAATLH
VREPVLSRLDPSREYVHACDPRENKLLGPLPFPHPHITHTVPSEVLLKTAQGRLF
YNTTLRKTYAADRKIINEIADPVVVGDNRLSLVSARLACIPIAIAANATWSPARR
RFPDPDPTFRFFGVRPVSILYRLRYLFALYCLPLNWLRRKHGSLSLGWDLCRIPT
DGDYTLADVPELVTYTNLPANHYRLGPVLMSPDKPPTWWHSLPTDRPIIYATIGSS
GGKTLLOVNLALADLPVITVIAAGRNHLKNVPANAFVADLPGEAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSNMDQHLNMEALERAGAVLLRTERLNTGEGVAAAQVQ
LSGAEFRQARRLPKPSDQTLPSRSTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translation="MVQTKRYAGTAAANTKKVAAWAPMFSIIIFTLNVAAYLPACLDS
IARCTGDFELVLDGGSTDETDIANIFAPNLGERLIIRHDTQGVVDANRWGVDLA
TGLVFLGADDSLYEARVAFTGEHPSDLVYGDVIMRSTNFRWGCAGFOLDRL
LFKENIQAIFVYERGLFGTIGPVNARVRLADWDNENRCSFNALVRYHVVVAVSY
NEFGGLSNTIVDEKFLKRLPMSTLGRILVILVVRPKVISRAMWRTVLSWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"

CDS
VAKIKEIVCDVDPGIDSAATATS"
908..1540
/codon_start=1
/transl_table=11
/product="u0002b"
/protein_id="AAAS0949.1"
/db_xref="GI:560528"
/translation="MVAARPAERSGDPAARVPVPSAWVVLIGGVIGLFPASMTLTVK
VRLDPTVPSNVPNTVSCSVNTTQASLLGPNPLGAGTFTVVVTVGLAVAK
VPLRWYVIGLAVGVAFVHMLIFQSLRIGALCPYCMVMAVIAATLVASIV
FGPRENGSERVCARLLYQWRWSLATLMTFTTFLIMVRFWDYNSTLI"
1565..4939
/codon_start=1
/transl_table=11
/product="pvc"
/protein_id="AAAS0948.1"
/db_xref="GI:560527"
/translation="MFSKVLVANRGEIAIRAFRAAYELGVGTAVVYVEDRNSQHLK
ADESQIGDIGHVPHAYLSVDEIVATARRAGADAIYFGYFELSENPDIAAACAAGIS
FVGPSEVLELAGNKSRAIAAREAGLFLMSAFVSDELLSVAAAGMPFLFYKAV
AGGGGRMVRVGDIAALPEAIEAASREAESAFGDPVYLEQAVINPHIEVQILADNL
GDVHLXERDCSVORRHOKVIELAPAPHLDAELRYKCMVDAAVAFARHIGYSCAGTVEF
LIDEGEVYFLEMPPRQVETHTVEITDVLVASQLRIAAGETLEQLGLQEDIAFH
GAAQCRIITDPANGFRPRAGSARCDPPAPVPSAWTAAPTWRKNQFVRLRHAGQAD
LSGRDLTPASRRARIAEFRIKGVSTNIPFLQAVLDDPDPFRAGVTTSPIDERPQLL
TARASDRGTIKLNPADVTNNPYGSRPTIYDDKLDPDLRAAPPAGSKQRLVKL
GPEGFAWLRESAAVGVDTTFRDAHQSLLATRVTSGLSVAPYLAKMTQQLLSVEC
WGAATYDVALRFLKEDPWERLATRAAMNCLQMLLRGNTVGTVPYPIVTSAFVQ
EATATGIDIFRIFDALNIESMRPAIDAVRETSGEIAEVAMCYGDLTDPEGEQIYTL
YTLKLEQIVDAGHVLAIKDMAGLLRPPAQRVLSALSRPDLPHVHTDHTFGQL
ASYVAWHAGADAVGAAPLAGTTSQALSIVAAAHTEYDGLSLVSALEAPYV
EALRYVAFESGLPGPTGVYHHEIPGQSLNFKQAIALGLGDRFEEIEEAVAGAD
RVLGRLVPTKVVGDALALVAGAVSDEFADPARFGIPESVLEFLREGLGDDP
GWPPEPLTALAGARPTAQLAADDEIALSSVGAKRQATLNEILLPPSPKSENEH
REAVGDTSQLANQFPYGLRQGEHVRKLERGVELLIGLEAISPEDEGRMTVMCILN
QOLRPVLRDTSIASAVPAEAKRGNFCHTAAAPPAGVTVGVGVGVAGVAGQTIATI
EAMKMEAPITAPAVAGTVERVAVSDTAEWRAETCWWV"
5057..5704
/codon_start=1
/transl_table=11
/product="u0002d"
/protein_id="AAAS0947.1"
/db_xref="GI:560526"
/translation="MQGYRPLAGDEAMRWGRYPLARESGLDPPHRRRRRIAYP
PRGTRPTDVRNGLSFLVARRDLTGLAVLDLVAGSGALGLEALSRGAASVLFVESD
QRSAVIARNTREALGLSATLRGCAVAVAAGTTSVDVLADPPVNVDSADVDAIL
AALGTNGTRGTAVVERATCAPLTWPEGRRWRPQRYVYGDTRLELAERLFANV"
5790..6275
/codon_start=1
/transl_table=11
/product="u0002e"
/protein_id="AAAS0946.1"
/db_xref="GI:560525"
/translation="MTGAVCPGSGFDPVTGLHVDIIFERAAAQFDEVVAILVNPAGTGM
FDLHERIAVMSVETPLNLRVQVHGLVDFVRSCTGTAIVKGLTGCTDPEYELQWA
QMNKHIAGVDTFFVATADRYSFVSSSLAKEVAMLGDDVSELLPEPVNRRRLDRLNT
T"
complement(7144..8076)
/codon_start=1
/transl_table=11
/product="purU"
/protein_id="AAAS0945.1"
/db_xref="GI:560524"
/translation="MKGSMTHAHATPNEPDYPPPPGPPPPADIGRLLLRCHDRPGII
AAVSTFLARAGANIISLDQHSAPEGGTFLOAIFHLPGLTAAVDELQDRDPSVADK
FGIDYRFAEAKRVAIMASTEDHCLDLRLNRNREGLEMSVVMVIANHPDLAAHVR
PFGVPFTHIPATRDTRAEQRLQSLGNDVLVLYARMQILSPGLEAIGCPLINI
HHSFLPAPTGAAPYORAREGVKLIATATVYVTEVLDBEGPIIEQDVVRVDHTHTVDLL
VRVGVADVERAVLSRAVLWHCODRVIVHNNQIVF"
complement(8149..9369)
/codon_start=1
/transl_table=11
/product="u0002f"
/protein_id="AAAS0944.1"
/db_xref="GI:560523"
/translation="MTSTKVEDRVTAAVLGAIGHALALTASMTWEILWALILGFALSA
VVOAVRRESTITVLLGDDRPRTLATIATGLGAASSCSVAVALASLFRKGANFTAAAM
AFEGISTNLVVELGIILALLMGWQFTAAEFVGGPMILVLAVFLFLFGVARGALIDARE
QAERGLAGSMEGHAAMDSIKREGSFWRRLISPPGFTSIHAHVFVMEWLAILRDLILGL
LIAGALAAWPESPFQWFFLANHPAWSWVGPIIGPIVAIVSFVCSIGNVPLAAVLWN
GGIFGGVIAFIFADLLIPLINITYKYTGAMMLVLGLTGYASVMVAGYITELLEFGT
TNLISQORSATVMTABEISWNTTWNVIFLVAALVVRFTTSGGLPMLRMWGGSPDA
PHLHDDHDDHDLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AAAS0943.1"
/db_xref="GI:560522"
/translation="MRASRLPALLPSLGPPLWLSRSALHHIPGGIYTVGVGNDRVAV
PLNVSPHSCGFPFRVSVCIYATASRGGPPVASEVRGDAISTTPDAAPGLAARRR
RILFVAEAVTLAHVVRFPALAQSLDPSRYEHPACDPYNNQLLGLPLPRHAIHTIPS
ERFENGLTQGRFYAMRTLKVEADRLVDEIAPDLVVGDLRISLSVSRHLAGIPIA
IANAWSPGAORRPLPDVITWTRLFGVRLVKLLYRLERPLLPALQCMPLNWWRRHGL
SSLCWNLCRIETDGDHTLYADVPELMTYDLPANHEVILGPVLMSPAGPPTPWDSLPT
DRPIVYATLGTSGGRNLLQLVLNALAELFVTVIAATAAATV"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AAAS0942.1"
/db_xref="GI:560521"
/translation="MPANAFVADYLPGEAAAARSADVVCNGGSLTTOQALVAGVPVIG
VAGNLDQHLNMEAVRACAGVLLRTERLKSQRVAGVMQVISRSEYPTSRERATRCLR
TRPWIIPAAECGCEPHAGEPAPDMVGQLVTST"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AAAS0941.1"
/db_xref="GI:560520"
/translation="MGLVRSRTSLVGLIGLVLVASFPAQLFYRPSDAVAEEVHKW
YVGNLVTKTYMGINCWKSVDMMNYQELISELQPSLVIEFGTRYGGSVAVYANIMR
QIGQFPKVLVDNSHKALDPPARREPDLFVSESSSTDPAIAEQRLKNEYFGKIFAI
LDSHSMNVHLAEMKLLRPLLSAGDYLVEDSSINGHNGHVPFGFGPGPYEAIEYBDEF
PNDYKDAERENKFGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AAAS0940.1"
/db_xref="GI:560519"
/translation="MEETSVAGDPDPAGTSTAPNAPEPVARRQRIILFVGEAATLH
VREPVLSRLDPSREYVHACDPRENKLLGPLPFPHPHITHTVPSEVLLKTAQGRLF
YNTTLRKTYAADRKIINEIADPVVVGDNRLSLVSARLACIPIAIAANATWSPARR
RFPDPDPTFRFFGVRPVSILYRLRYLFALYCLPLNWLRRKHGSLSLGWDLCRIPT
DGDYTLADVPELVTYTNLPANHYRLGPVLMSPDKPPTWWHSLPTDRPIIYATIGSS
GGKTLLOVNLALADLPVITVIAAGRNHLKNVPANAFVADLPGEAAARLAVVLCN
GGSPTQOALAAAGVPVIGLPSNMDQHLNMEALERAGAVLLRTERLNTGEGVAAAQVQ
LSGAEFRQARRLPKPSDQTLPSRSTSKVRCGSSAKGCPYRRLANGCPVEHLIRER
VAVNAEVPQRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kc"
/protein_id="AAAS0938.1"
/db_xref="GI:560517"
/translation="MVQTKRYAGTAAANTKKVAAWAPMFSIIIFTLNVAAYLPACLDS
IARCTGDFELVLDGGSTDETDIANIFAPNLGERLIIRHDTQGVVDANRWGVDLA
TGLVFLGADDSLYEARVAFTGEHPSDLVYGDVIMRSTNFRWGCAGFOLDRL
LFKENIQAIFVYERGLFGTIGPVNARVRLADWDNENRCSFNALVRYHVVVAVSY
NEFGGLSNTIVDEKFLKRLPMSTLGRILVILVVRPKVISRAMWRTVLSWRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"

```
/protein_id="AAA50939.1"
/db_xref="GI:560518"
/translation="NKSILARFTARSAAFEVSRYSERDLKHQFVKOLKSRVDVVF
DVANGGYAAGLRRAAYKGRIVPEPLSGPFTILESASTDPLWDCKQHAGSDGT
VTINTAGAGSSSVLPMLKSHQNAFPFANYGTQEASIHRLDSVAPEFLGMGVAF
KVDVQGFQKVLGAGKSTIDHCVGMQLELSFLPLYEGGMLIPEALDIVSLGFTLTG
LLPCFIDANNRMLQADGIFPREDD"
CDS
17208..17933
/codon_start=1
/transl_table=11
/product="u00021"
/protein_id="AAA50937.1"
/db_xref="GI:560516"

Query Match      43.2%; Score 346; DB 1; Length 31175;
Best Local Similarity 71.0%; Pred. No. 3.5e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60
Db 15076 ATGGCGGACCAATGTTTTCGATCATATCCCACTTGAACTGGCTGCGGTATTGCT 15017
QY 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAAAGTGGTCTCTTGTGAC 120
Db 15016 GCCTGCCTCGACAGCATCGCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180
Db 14956 GCGGCTCGACGAGCAAAACCTCGACATCGCAACATTTTTCGCCCCCAACCTCGGCGAG 14897
QY 181 GCACTGGTGTCTCACAGCGGGCCCGATGATGCGCCCTACGAGCGCCATGAACCGCGGGCTC 240
Db 14896 CGGTTGATCATTCATCGGACACCGACCGAGCGGCTCTACGAGCCATGAACCGCGGGCTG 14837
QY 241 GCGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGAGCAGACACCTCTACGAAACCA 300
Db 14836 GACCTGGCCACGGAACGTGGTGTCTTTCTGGCGCGGAGCAGACGCTGTACGAGGCT 14777
QY 301 ACACGTTGGCCAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 14776 GACACCTGGCGGGTGGCGGCTTCATTGGCGAACAACGAGCCAGCGATCTGGTATAT 14717
QY 361 GCGGATGTGTGATCGTTCGACGAAAGCCGACATGCGGACCTTCGACCTCGACCGC 420
Db 14716 GCGGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCCAATCGATCTTTTACCGCGCGTGAGCTTTTCGAC 480
Db 14656 CTGTTGTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTCGGC 14597
QY 481 GGCATCGGCCCTTACAACCTGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGC 540
Db 14596 ACCATCGGTCCCTACAACCTCGCTACCGGGTCTTGGCGGACTGGGACTTCAATATTGC 14537
QY 541 TGCTTCTCAACCGCGCTGATTACCGCTACATGGAGCTCGTGAATTCGGAATACAA 600
Db 14536 TGCTTTTCAACCGAGCGCTCGTCAACCGCTACATGCACTGGTGTGTTGCAAGTACAA 14477
QY 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 14476 GAATTCGGGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCGG 14420
QY 661 ATGTAC 666
Db 14419 ATGTCC 14414
```

```
RESULT 12
MSGY2/c      MSGY2      31176 bp      DNA      linear      BCT 03-DEC-1996
LOCUS      Mycobacterium tuberculosis sequence from clone y2.
DEFINITION      AD000009
ACCESSION      AD000009.1
VERSION      GI:1702967
```

```
KEYWORDS      . Mycobacterium tuberculosis
SOURCE      Mycobacterium tuberculosis
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
      Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
      tuberculosis complex.
REFERENCE      1 (bases 1 to 31176)
AUTHORS      Du, L.
TITLE      Direct Submission
JOURNAL      Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
      Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT      GSDb:S:1004709.
FEATURES
      Location/Qualifiers
      source      1..31176
      /organism="Mycobacterium tuberculosis"
      /mol_type="genomic DNA"
      /db_xref="taxon:1773"
      /clone="y2"

ORIGIN
Query Match      43.2%; Score 346; DB 1; Length 31176;
Best Local Similarity 71.0%; Pred. No. 3.5e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60
Db 15076 ATGGCGGACCAATGTTTTCGATCATATCCCACTTGAACTGGCTGCGGTATTGCT 15017
QY 61 GCCTGCCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGGAAAGTGGTCTCTTGTGAC 120
Db 15016 GCCTGCCTCGACAGCATCGCCGTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTCGCGCCCGGAACCTCGGCTCG 180
Db 14956 GCGGCTCGACGAGCAAAACCTCGACATCGCAACATTTTTCGCCCCCAACCTCGGCGAG 14897
QY 181 CGACTGGTGTCTCACAGCGGGCCCGATGATGCGCCCTACGAGCGCCATGAACCGCGGGCTC 240
Db 14896 CGGTTGATCATTCATCGGACACCGACCGAGCGGCTCTACGAGCCATGAACCGCGGGCTG 14837
QY 241 GCGTAGCCACAGGCGAATGGGTACTTTTTTAGCGCGCGAGCAGACACCTCTACGAAACCA 300
Db 14836 GACCTGGCCACGGAACGTGGTGTCTTTCTGGCGCGGAGCAGACGCTGTACGAGGCT 14777
QY 301 ACACGTTGGCCAGGTAGCGCTTTTCTCGGCGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 14776 GACACCTGGCGGGTGGCGGCTTCATTGGCGAACAACGAGCCAGCGATCTGGTATAT 14717
QY 361 GCGGATGTGTGATCGTTCGACGAAAGCCGACATGCGGACCTTCGACCTCGACCGC 420
Db 14716 GCGGACGTGATCATGCGCTCAACCAATTCGCTGGGGTGGGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCCAATCGATCTTTTACCGCGCGTGAGCTTTTCGAC 480
Db 14656 CTGTTGTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTCGGC 14597
QY 481 GGCATCGGCCCTTACAACCTGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGC 540
Db 14596 ACCATCGGTCCCTACAACCTCGCTACCGGGTCTTGGCGGACTGGGACTTCAATATTGC 14537
QY 541 TGCTTCTCAACCGCGCTGATTACCGCTACATGGAGCTCGTGAATTCGGAATACAA 600
Db 14536 TGCTTTTCAACCGAGCGCTCGTCAACCGCTACATGCACTGGTGTGTTGCAAGTACAA 14477
QY 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 14476 GAATTCGGGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCGG 14420
QY 661 ATGTAC 666
Db 14419 ATGTCC 14414
```

```
RESULT 13
AR345347/c
LOCUS AR345347 3215 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 1 from patent US 6583266.
ACCESSION AR345347
VERSION AR345347.1 GI:33742001
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3215)
AUTHORS Smith,D.R. and Mao,J.-i.
TITLE Nucleic acid and amino acid sequences relating to mycobacterium tuberculosis and leprae for diagnostics and therapeutics
JOURNAL Patent: US 6583266-A 1 24-JUN-2003;
FEATURES
    Location/Qualifiers
        1..3215
            /organism="unknown"
            /mol_type="genomic DNA"
ORIGIN
    Query Match 43.2%; Score 346; DB 6; Length 32155;
    Best Local Similarity 71.0%; Pred. No. 3.5e-63;
    Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 1 ATGACTGGCCAGTGTCTCGATAATTATCCTTACCTTCAATGACGGGTGACGCTGCAA 60
DB 15076 ATGGCGGCACCAATGTTTCGATCATCCCACTTGAAGTGGCTGGGTATTGCT 15017
QY 61 GCCTGCTCGGAAGCATGTCGGGACACCTACCGGGAAGTGAAGTGGTCTTGTGAC 120
DB 15016 GCGTCGCTCGACAGCATGCCCGTCCAGACCTCGGTGACTTCGAGCTGTGTCGAC 14957
QY 121 GCGGTTTCGACCGATCGGACCTCGACATCGGAACAGTTTCGCCCGCGAACTCGGCTCG 180
DB 14956 GCGGCTCGACGGAACCACTTCGATCGCAATCTTCGCCCGCACTTCGCGGAG 14897
QY 181 CACTGTGTGTTCAAGCGGCGCCGATGATGCCCCCTACGACCGCATGAACCGGCGCTC 240
DB 14896 CGTTTGTATCATTCGCGACACCGACCGACCGGCGTCTACGACCGCATGAACCGGCGCTG 14837
QY 241 GCGTAGGACACAGGGAATGGTACTTTTATGCGGCGGACGACACCTCTACGAACA 300
DB 14836 GACCTGGCGCACCGAAGTGGTGTCTTTCTGGGCGGCGGACGACGCTGTACGAGCT 14777
QY 301 ACCAGCTTGGCCAGGTAGCGCTTTCTCGCGACCATGCGGCAAGCCATCTTCTCTAT 360
DB 14776 GACACCTTGGCGGCGGTCGCCCTTCATGCGGAACAGACCGCGATCTGTATAT 14717
QY 361 GCGGATGTGTGATCGGTTTCGACGAAAGCCGCGATGCGGACCTTCGACCTCGACCGC 420
DB 14716 GCGGACGTGATCATCGCTCAACCAATTTCCGCTGGGCTGGCGCTTCGACCTCGACCGT 14657
QY 421 CTCCTATTGAGACGAATTTGTGCCACCATGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 14656 CTGTTGTTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTCGCG 14597
QY 481 GGCATCGGCTTTACACCTGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTCGC 540
DB 14596 ACCATCGGCTTCAACCTCGCTACCGGCTTGGCGGACTGGGACTTCAATATTCGC 14537
QY 541 TGCTTCTCAACCCCGGCGTGAATACCGCTACATGACGCTGCTGATTTCCGAATACAA 600
DB 14536 TGCTTTTCCAAACCGAGCTCGTCAACCGCTACATGACGCTGCTGATTTGGAAGTACAA 14477
QY 601 GACATGACCGCTTTCAGCATGAGGACGGGACTGATAAGAGTTTCAGAAAACGCTGCCA 660
DB 14476 GAAATCGGCGGCTTCAGCAATACGATCG---TCGACAAGGAGTTTTCGAGCGGCTGCG 14420
QY 661 ATGTAC 666
DB 14419 ATGTCC 14414
```

RESULT 14
AE000516.32
WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name

Begin End
AE000516.00 1 110000
AE000516.01 100001 210000
AE000516.02 200001 310000
AE000516.03 300001 410000
AE000516.04 400001 510000
AE000516.05 500001 610000
AE000516.06 600001 710000
AE000516.07 700001 810000
AE000516.08 800001 910000
AE000516.09 900001 1010000
AE000516.10 1000001 1110000
AE000516.11 1100001 1210000
AE000516.12 1200001 1310000
AE000516.13 1300001 1410000
AE000516.14 1400001 1510000
AE000516.15 1500001 1610000
AE000516.16 1600001 1710000
AE000516.17 1700001 1810000
AE000516.18 1800001 1910000
AE000516.19 1900001 2010000
AE000516.20 2000001 2110000
AE000516.21 2100001 2210000
AE000516.22 2200001 2310000
AE000516.23 2300001 2410000
AE000516.24 2400001 2510000
AE000516.25 2500001 2610000
AE000516.26 2600001 2710000
AE000516.27 2700001 2810000
AE000516.28 2800001 2910000
AE000516.29 2900001 3010000
AE000516.30 3000001 3110000
AE000516.31 3100001 3210000
AE000516.32 3200001 3310000
AE000516.33 3300001 3410000
AE000516.34 3400001 3510000
AE000516.35 3500001 3610000
AE000516.36 3600001 3710000
AE000516.37 3700001 3810000
AE000516.38 3800001 3910000
AE000516.39 3900001 4010000
AE000516.40 4000001 4110000
AE000516.41 4100001 4210000
AE000516.42 4200001 4310000
AE000516.43 4300001 4403837

Continuation (33 of 44) of AE000516 from base 3200001 (AE000516 Mycobacterium tuberculosis

Query Match 43.2%; Score 346; DB 1; Length 110000;

Best Local Similarity 71.0%; Pred. No. 3.6e-63;

Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGGCCAGTGTCTCGATAATTATCCTTACCTTCAATGACGGGTGACGCTGCAA 60

DB 103847 ATGGCGGCACCAATGTTTCGATCATCCCCACCTTGAACGCTGCGGTATTGCT 103906

QY 61 GCCTGCTCGGAACGATCGTCGGGACACCTTACCGGGAAGTGAAGTGGTCTTGTTCGAC 120

DB 103907 GCCTGCTCGACAGCATCGCCCGTCAGACCTGCGGTGACTTCGAGCTGTTGTTGTCGAC 103966

QY 121 GCGGTTTCGACCGATCGGAACCTTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCG 180

DB 103967 GCGGCTCGACGACGGAACCTTCGACATCGCAACATTTTCGCCCCCAACTCGCGCAG 104026

QY 181 CGACTGTTGTTTCACAGCGGCGCGATGATGGCCCTTACGACGCAATGAACGCGGCGCTC 240

DB 104027 CGGTTGATTCATTTCATCGGACACCGCAGGCGCTCTACGACGCAATGAACGCGGCGTG 104086

QY 241 GCGGTAGCCACAGGCGAATGGTACTTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300

```
Db 104087 GACCTGGCCACGGAAGTGGTTGCTCTTTCTGGCGCGGACGACAGCTGTACGAGGCT 1041146
Qy 301 ACCAGCTGGCCAGGTAGCCGCTTTCTCGCGGACCATGGCGGAAGCACTTGTCTAT 360
Db 104147 GACACCCCTGGCGCGGTGGCCGCTTCATTGGCGAACAACGAGCCGAGGATCTGGTATAT 104206
Qy 361 GCGATGTTGTGATCGTTTCGACGAAAGCGGCATGCGGACCTTCGACCTCGACCGC 420
Db 104207 GCGACGTGATCATGGCTCAACCAATTTCCGCTGGGTGGCGCTTGACCTCGACCGT 104266
Qy 421 CTCCTATTGTAGACGAATTTGSCCACCAAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 104267 CTGTTTTCAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTCGGC 104326
Qy 481 GGCATCGGCCCTTACAACTCGGCTACCGAGTCTGGCGGAGTGGGACTTCAATATTGCG 540
Db 104327 ACCATCGGTCCTACAACTCCGCTACCGGTCTCGCGGACTTGGGACTTCAATATTGCG 104386
Qy 541 TGCCTTCTCAACCCGCGCTGATTACCGCTACATGGAGCTGCTGATTTCCGAATACAAC 600
Db 104387 TGCCTTTTCAACCCAGCGCTGCTACCGCTACATGCACTGCTGCTTGCAGCTACAAC 104446
Qy 601 GACATGACCGGCTTCAGCATGAGGCGAGGAGTGTATAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 104447 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAGGAGTTTTTCGAGCGGCTGCGG 104503
Qy 661 ATGTAC 666
Db 104504 ATGTCC 104509
```

RESULT 15
AE000516_33
WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name	Begin	End
AE000516_00	1	110000
AE000516_01	100001	210000
AE000516_02	200001	310000
AE000516_03	300001	410000
AE000516_04	400001	510000
AE000516_05	500001	610000
AE000516_06	600001	710000
AE000516_07	700001	810000
AE000516_08	800001	910000
AE000516_09	900001	1010000
AE000516_10	1000001	1110000
AE000516_11	1100001	1210000
AE000516_12	1200001	1310000
AE000516_13	1300001	1410000
AE000516_14	1400001	1510000
AE000516_15	1500001	1610000
AE000516_16	1600001	1710000
AE000516_17	1700001	1810000
AE000516_18	1800001	1910000
AE000516_19	1900001	2010000
AE000516_20	2000001	2110000
AE000516_21	2100001	2210000
AE000516_22	2200001	2310000
AE000516_23	2300001	2410000
AE000516_24	2400001	2510000
AE000516_25	2500001	2610000
AE000516_26	2600001	2710000
AE000516_27	2700001	2810000
AE000516_28	2800001	2910000
AE000516_29	2900001	3010000
AE000516_30	3000001	3110000
AE000516_31	3100001	3210000
AE000516_32	3200001	3310000
AE000516_33	3300001	3410000
AE000516_34	3400001	3510000
AE000516_35	3500001	3610000

```
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4403837
Continuation (34 of 44) of AE000516 from base 3300001 (AE000516 Mycobacterium tuberculosis

Query Match 43.2%; Score 346; DB 1; Length 110000;
Best Local Similarity 71.0%; Pred. No. 3.6e-63;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy 1 ATGACTCGCCAGTGTCTCGATAATTATCCCTACCTCAATGACAGCGGTGACGCTGCAA 60
Db 3847 ATGCGCGCACCAATGTTTTTCGATCATATCCCCACCTTGAAACGTGGCTGCGGTATTGCT 3906
Qy 61 GCCTGCCTCGGAAGCATCGTCGGGCAGACCTTACCGGGAAGTGGAGTGGTCTTTGTTCGAC 120
Db 3907 GCCTGCCTCGACAGCATGCCCGCTCAGACCTGCGGTGACTTCGAGCTGGTACTGGTCGAC 3966
Qy 121 GCGGTTTCGACCGATCGGACCTTCGACATCGCAACAGTTTTCGCCCGCGGAACCTCGGTCG 180
Db 3967 GCGGCTCGACGGACGAAACCTTCGACATCGCAACATTTTCGCCCCCAACCTCGCGGAG 4026
Qy 181 CGACTGTCGTTACAGGGGGCCCGATGATGGCCCTTACGACGCATGAACCGGGGGGTC 240
Db 4027 CGGTTGATCATTTTCGCGACACCGAGGCGTCTACGACGCGCATGAACCGGGGGGTG 4086
Qy 241 GCGTAGCCACAGCGGAATGGGTACTTTTTTAGCGCGCGACACACACCTCTACGAACCA 300
Db 4087 GACCTGGCCACCAGAAAGCTGGTTGCTCTTTCGGGCGGAGACAGAGCTGTACGAGCT 4146
Qy 301 ACCAGTTGGCCAGGTAGCGGCTTTTCTCGGACCATGCGGAAGCCATCTTTGTCTAT 360
Db 4147 GACACCTGGCGGGTGGCGGCTTCATTGGCGAAACACGAGCCAGCGATCTGGTATAT 4206
Qy 361 GCGCATGTTGTAGCGTTCGACGAAAGCGGCATGCGCGACCTTTCGACCTCGACCGC 420
Db 4207 GCGACGCTGATCATGCGCTCAACCAATTTCCGCTGGGGTGGCGCTTTCGACCTCGACCGT 4266
Qy 421 CTCCTATTTCGACGAAATTTGTGCCACCAATTCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
Db 4267 CTGTTGTTCAAGCGCAACATCTGCCATCAGCGGATCTTACCGCGCGGAGCTCTTCGGC 4326
Qy 481 GGCATCGGCCCTTACAACTGCGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
Db 4327 ACCATCGGTCCTACAACTCCGCTACCGGTCCTGCGCGGACTGGGACTTCAATATTGCG 4386
Qy 541 TGCTTCTCAACCGCGGCTGATTTACCGCTACATGGAGCTGCTGATTTCCGAATACAAC 600
Db 4387 TGCTTTTCCAAACCCAGCGCTCGTCCCGGCTACATGCACTGCGTGGTCTTTCGAAGCTACAAC 4446
Qy 601 GACATGACCGGCTTCAGCATGAGGCGAGGACTGTATAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 4447 GAATTCGGCGGCTCAGCAATACGATCG---TCGACAGGAGTTTTTCGAGCGGCTGCGG 4503
Qy 661 ATGTAC 666
Db 4504 ATGTCC 4509
```

Search completed: March 6, 2005, 21:05:08
Job time : 3865 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2005, 21:05:14 ; Search time 597 Seconds
(without alignments)
7961.493 Million cell updates/sec.

Title: US-10-805-311-23

Perfect score: 801

Sequence: 1 atgactgcccagtggtctc.....aacgaagcgcagaaccgtag 801

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5401638 seqs, 2966923429 residues

Total number of hits satisfying chosen parameters: 10803276

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09D_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	801	100.0	801	18	US-10-805-311-23
2	801	100.0	4435	18	US-10-805-311-4
3	799.4	99.8	801	18	US-10-805-311-21
4	799.4	99.8	7995	18	US-10-805-311-3
5	346	43.2	768	17	US-10-282-122A-26361
6	346	43.2	828	17	US-10-282-122A-28624
7	346	43.2	828	18	US-10-805-311-38
8	69.2	8.6	2334	15	US-10-156-761-5079
9	69.2	8.6	9025608	15	US-10-156-761-1
10	67.4	8.4	2196	15	US-10-156-761-5077
11	60	7.5	900	17	US-10-282-122A-30023

12	57	7.1	2256646	18	US-10-470-565-1	Sequence 1, Appli
13	55.6	6.9	1060	15	US-10-193-002-306	Sequence 306, App
14	55.6	6.9	1060	15	US-10-084-843-311	Sequence 311, App
15	55.4	6.9	1033	17	US-10-282-122A-36601	Sequence 36601, A
16	53.8	6.7	1035	17	US-10-282-122A-39197	Sequence 39197, A
17	53.4	6.7	4512	15	US-10-156-761-2222	Sequence 2222, Ap
18	48	6.0	738	17	US-10-424-599-107538	Sequence 107538,
19	47.8	6.0	1022	15	US-10-193-002-325	Sequence 325, App
20	47.8	6.0	1022	15	US-10-084-843-330	Sequence 330, App
21	45.2	5.6	1827	17	US-10-108-260A-139	Sequence 139, App
22	42.8	5.3	753	15	US-10-156-761-6137	Sequence 6137, Ap
23	42.8	5.3	4725	15	US-10-205-032-17	Sequence 17, Appli
24	42.8	5.3	5859	13	US-10-007-267-1	GENERAL INFORMA
25	42.8	5.3	5859	13	US-10-007-267-7	GENERAL INFORMA
26	42.8	5.3	5859	17	US-10-096-129-1	Sequence 1, Appli
27	42.8	5.3	5859	17	US-10-654-528-1	Sequence 1, Appli
28	42.8	5.3	5859	17	US-10-654-528-7	Sequence 7, Appli
29	42.8	5.3	60196	15	US-10-205-032-1	Sequence 1, Appli
30	42.8	5.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
31	42.6	5.3	1326	17	US-10-282-122A-23449	Sequence 23449, A
32	42.4	5.3	1139	18	US-10-425-115-78950	Sequence 78950, A
33	42.4	5.3	1143	15	US-10-156-761-823	Sequence 823, App
34	42.2	5.3	2664	18	US-10-437-963-61100	Sequence 61100, A
35	41.4	5.2	522	17	US-10-424-599-73486	Sequence 73486, A
36	40.8	5.1	320	16	US-10-259-678-467	Sequence 467, App
37	40.8	5.1	882	17	US-10-282-122A-33586	Sequence 33586, A
38	40.8	5.1	1836	10	US-09-953-348-52	Sequence 52, Appl
39	40.8	5.1	1836	15	US-10-267-255-52	Sequence 52, Appl
40	40.8	5.1	53500	10	US-09-953-348-76	Sequence 76, Appl
41	40.8	5.1	53500	15	US-10-267-255-76	Sequence 76, Appl
42	40.6	5.1	2073	17	US-10-282-122A-11384	Sequence 11384, A
43	40.6	5.1	2427	18	US-10-437-963-69728	Sequence 69728, A
44	40.4	5.0	1443	15	US-10-156-761-4460	Sequence 4460, Ap
45	40.4	5.0	2811	10	US-09-953-348-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-10-805-311-23
; Sequence 23, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Mycobacterium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(798)

US-10-805-311-23

```
Query Match      100.0%; Score 801; DB 18; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.6e-255;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
DB 1 ATGACTGGCCAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60

QY 61 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGGTCTTGTGAC 120
DB 61 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGGTCTTGTGAC 120

QY 121 GCGGTTTCGACCGATCGACCTCGACATCGGAACAGTTCGCGCCGGAACCTGGCTCG 180
DB 121 GCGGTTTCGACCGATCGACCTCGACATCGGAACAGTTCGCGCCGGAACCTGGCTCG 180

QY 181 CGACTGGTCTGTTACAGCGGGCCCGATGATGSCCCCTACGACGCCATGAACCGGGCGTC 240
DB 181 CGACTGGTCTGTTACAGCGGGCCCGATGATGSCCCCTACGACGCCATGAACCGGGCGTC 240

QY 241 GCGTAGCCACAGGCGAATGGGTACTTTTTAGGCGCGACGACACACCTCTACGAACCA 300
DB 241 GCGTAGCCACAGGCGAATGGGTACTTTTTAGGCGCGACGACACACCTCTACGAACCA 300

QY 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGAGCCATGCGGCAAGCCATCTTGTCTAT 360
DB 301 ACCAGTTGGCCAGGTAGCGCTTTCTCGGAGCCATGCGGCAAGCCATCTTGTCTAT 360

QY 361 GCGGATGTTGTATGCGTTTCAGCAAGAAAGCGGACCTGCGGACCTTCGACCTCGACCGC 420
DB 361 GCGGATGTTGTATGCGTTTCAGCAAGAAAGCGGACCTGCGGACCTTCGACCTCGACCGC 420

QY 421 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 421 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480

QY 481 GGCATCGGCCCTTACAACTGGGCTACCGAGTCTGGGGGACCTGGGACTTCAATATTGCG 540
DB 481 GGCATCGGCCCTTACAACTGGGCTACCGAGTCTGGGGGACCTGGGACTTCAATATTGCG 540

QY 541 TGCCTTCTCAACCCGGCGCTGATTACCGCTACATGGAAGTCTGATTTCCGAATACAAAC 600
DB 541 TGCCTTCTCAACCCGGCGCTGATTACCGCTACATGGAAGTCTGATTTCCGAATACAAAC 600

QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGGAATGATAAAGAGTTTCAGAAAACGGCTGCCA 660
DB 601 GACATGACCGGCTTCAGCATGAGGCGAGGGAATGATAAAGAGTTTCAGAAAACGGCTGCCA 660

QY 661 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGCATGCTGGCGTTTGTGAAGAC 720
DB 661 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGCATGCTGGCGTTTGTGAAGAC 720

QY 721 AAGGAGAAATCGCGCTTCGCTTGGTACCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
DB 721 AAGGAGAAATCGCGCTTCGCTTGGTACCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780

QY 781 GAACGAAGCGCAGAACCGGTAG 801
DB 781 GAACGAAGCGCAGAACCGGTAG 801
```

RESULT 2

```
US-10-805-311-4
; Sequence 4, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
```

```
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC
; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4435
; TYPE: DNA
; ORGANISM: Mycobacterium
US-10-805-311-4

Query Match      100.0%; Score 801; DB 18; Length 4435;
Best Local Similarity 100.0%; Pred. No. 1.4e-254;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGGCCAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 60
DB 3335 ATGACTGGCCAGTGTCTCGATAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAA 3394

QY 61 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGGTCTTGTGAC 120
DB 3395 GCCTGCTCGGAAGCATCGTCGGGAGACCTACCGGGAAGTGAAGTGGTCTTGTGAC 3454

QY 121 GCGGTTTCGACCGATCGGACCTCTCGACATCGCAACAGTTCGCGCCGGAACCTGGCTCG 180
DB 3455 GCGGTTTCGACCGATCGGACCTCTCGACATCGCAACAGTTCGCGCCGGAACCTGGCTCG 3514

QY 181 CGACTGGTCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCCATGAACCGGGCGTC 240
DB 3515 CGACTGGTCTGTTACAGCGGGCCCGATGATGCGCCCTACGACGCCATGAACCGGGCGTC 3574

QY 241 GCGGTAGCCACAGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300
DB 3575 GCGGTAGCCACAGCGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA 3634

QY 301 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCGAACAATGCGGCAAGCCATCTTGTCTAT 360
DB 3635 ACCAGTTGGCCAGGTAGCGCTTTTCTCGGCGAACAATGCGGCAAGCCATCTTGTCTAT 3694

QY 361 GCGGATGTTGTATGCGTTTCAGCAAGAAAGCGGCAATGCGGACCTTCGACCTCGACCGC 420
DB 3695 GCGGATGTTGTATGCGTTTCAGCAAGAAAGCGGCAATGCGGACCTTCGACCTCGACCGC 3754

QY 421 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
DB 3755 CTCCTATTGAGACGAATTTGTGCCAACAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 3814

QY 481 GGCATCGGCCCTTACAACTGGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 540
DB 3815 GGCATCGGCCCTTACAACTGGGCTACCGAGTCTGGGCGGACTGGGACTTCAATATTGCG 3874

QY 541 TGCTTCTCCAAACCGCGCGCTGATTACCGCTACATGGAAGTTCGGAATACAAAC 600
DB 3875 TGCTTCTCCAAACCGCGCGCTGATTACCGCTACATGGAAGTTCGGAATACAAAC 3934

QY 601 GACATGACCGGCTTCAGCATGAGGCGAGGGAATGATAAAGAGTTTCAGAAAACGGCTGCCA 660
DB 3935 GACATGACCGGCTTCAGCATGAGGCGAGGGAATGATAAAGAGTTTCAGAAAACGGCTGCCA 3994

QY 661 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGCATGCTGGCGTTTGTGAAGAC 720
DB 720 ATGTACTTCTGGGTTGCAAGGTGGGAGACTTTCAGGCGCATGCTGGCGTTTGTGAAGAC 720
```


Db 3995 ATGACTTCTGGGTTGCGAGGTGGAGACTTCGAGCGGCATGCTGGCGTTTGTGAAGAC 4054
Qy 721 AAGGAGATCGCGTCTGCGCTTGGGTACGCGGTTGATAAGGGTTAAGCCCGTCTCCAAA 780
Db 4055 AAGGAGATCGCGTCTGCGCTTGGGTACGCGGTTGATAAGGGTTAAGCCCGTCTCCAAA 4114
Qy 781 GAACGAAGCGGAGAACCGTAG 801
Db 4115 GAACGAAGCGGAGAACCGTAG 4135

RESULT 3

US-10-805-311-21
; Sequence 21, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC
; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Mycobacterium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(798)
US-10-805-311-21

Query Match 99.8%; Score 799.4; DB 18; Length 801;

Best Local Similarity 99.9%; Pred. No. 2.6e-254;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Db 1 ATGACTGCGCCAGTGTCTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60
Qy 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGGAAGTGGAGTGTCTTGTTCGAC 120
Db 61 GCCTGCTCGGAAGCATCGTGGGCGAGACCTACCGGGGAAGTGGAGTGTCTTGTTCGAC 120
Qy 121 GCGGTTTCGACGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Db 121 GCGGTTTCGACGATCGGACCTCGACATCGGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
Qy 181 CGACTGTTGTTTCACAGCGGGCCCGCATGATGSCCCCTACGAGCCCATGAACCGGCGGTC 240
Db 181 CGACTGTTGTTTCACAGCGGGCCCGCATGATGSCCCCTACGAGCCCATGAACCGGCGGTC 240
Qy 241 GCGGTAGCCACAGGCGGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300
Db 241 GCGGTAGCCACAGGCGGAATGGGTACTTTTTTAGCGCGCGACGACACCTCTACGAACCA 300
Qy 301 ACCAGTTGGCCAGGTAGCGCGCTTTCTCGCGGACCATGCGGCAAGCCATCTTGTCTAT 360
Db 301 ACCAGTTGGCCAGGTAGCGCGCTTTCTCGCGGACCATGCGGCAAGCCATCTTGTCTAT 360

Db 301 ACCAGTTGGCCAGGTAGCGCGCTTTCTCGCGGACCATGCGGCAAGCCATCTTGTCTAT 360
Qy 361 GCGGATGTTGTATGCTGCTTCGACGAAAAAGCGGCATGCGGACCTTTTCGACCTCGACCGC 420
Db 361 GCGGATGTTGTATGCTGCTTCGACGAAAAAGCGGCATGCGGACCTTTTCGACCTCGACCGC 420
Qy 421 CTCTTATTTGAGACGAATTTGTGTGCCACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC 480
Db 421 CTCTTATTTGAGACGAATTTGTGTGCCACCAATCGATCTTTTACCGCGTGAGCTTTTCGAC 480
Qy 481 GGCATCGGCCCTTACACCTGCGTACCGAGTCTGGGGGGGACTGGGACTTCAATATTCGC 540
Db 481 GGCATCGGCCCTTACACCTGCGTACCGAGTCTGGGGGGGACTGGGACTTCAATATTCGC 540
Qy 541 TGCTTCTCCAAACCGCGCGCTGATTACCGCTACATGACGCTCGTGAATTTCCGAAATACAAC 600
Db 541 TGCTTCTCCAAACCGCGCGCTGATTACCGCTACATGACGCTCGTGAATTTCCGAAATACAAC 600
Qy 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Db 601 GACATGACCGGCTTCAGCATGAGGAGGGGACTGATAAAGAGTTTCAGAAAAACGGCTGCCA 660
Qy 661 ATGCTACTTCTGGTTCAGGCTGGAGACTTGCAGGCGCATGCTGGCGTTTGTGAAGAC 720
Db 661 ATGCTACTTCTGGTTCAGGCTGGAGACTTGCAGGCGCATGCTGGCGTTTGTGAAGAC 720
Qy 721 AAGGAGATCGCGCTTGGCTTACGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
Db 721 AAGGAGATCGCGCTTGGCTTACGCGGTTGATAAGGGTTAAGCGCGTCTCCAAA 780
Qy 781 GAACGAAGCGGAGAACCGTAG 801
Db 781 GAACGAAGCGGAGAACCGTAG 801

RESULT 4

US-10-805-311-3
; Sequence 3, Application US/10805311
; Publication No. US20040260078A1
; GENERAL INFORMATION:
; APPLICANT: Hermon-Taylor, John
; APPLICANT: Doran, Tim
; APPLICANT: Millar, Douglas
; APPLICANT: Tizard, Mark
; APPLICANT: Loughlin, Mark
; APPLICANT: Sumar, Nazira
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC
; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND
; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY
; FILE REFERENCE: 117-260
; CURRENT APPLICATION NUMBER: US/10/805,311
; CURRENT FILING DATE: 2004-03-22
; PRIOR APPLICATION NUMBER: US/09/705,911
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: US/09/091,538
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/GB96/03221
; PRIOR FILING DATE: 1996-12-23
; PRIOR APPLICATION NUMBER: GB 9526178.0
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7995
; TYPE: DNA
; ORGANISM: Mycobacterium
US-10-805-311-3

Query Match 99.8%; Score 799.4; DB 18; Length 7995;

Best Local Similarity 99.9%; Pred. No. 5.8e-254;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGACTGCGCAGTGTCTTCGATAATTATCCCTACCTTCAATGACGCGGTGACGCTGCAA 60

4947	Db		ATGACTGCGGCAGTGTCTCGATAAATATCCCTACCTTCAATGACAGCGGTGACGCTGCAAA	5006
61	Qy		GCCTGCGCTCGGAAGCATCGTCGGGCAGACCTTACCGGGAAGTGGGAAGTGGCTTTGTCTGCAC	120
5007	Db		GCCTGCGCTCGGAAGCATCGTCGGGCAGACCTTACCGGGAAGTGGGAAGTGGCTTTGTCTGCAC	5066
121	Qy		GGCGGTTTCGACCGATCGGAACCTCGACATCGCGAAACAGTTTTCGCGCCGGGAACTCGGCTCG	180
5067	Db		GGCGGTTTCGACCGATCGGAACCTCGACATCGCGAAACAGTTTTCGCGCCGGGAACTCGGCTCG	5126
181	Qy		CGACTGTGTGTTACAGCGGCGCCGATGATGGCCCTTACGACGCCATGAAACCGCGGGGTC	240
5127	Db		CGACTGTGTGTTACAGCGGCGCCGATGATGGCCCTTACGACGCCATGAAACCGCGGGGTC	5186
241	Qy		GGCGTAGCCACAGCGGAATGGGTACTTTTTTATAGGCGCGACGACACCCCTCTAGAACCA	300
5187	Db		GGCGTAGCCACAGCGGAATGGGTACTTTTTTATAGGCGCGCGACGACACCCCTCTAGAACCA	5246
301	Qy		ACCAAGTTGGCCAGGTAGCGCTTTTCTCGGCGAACCATCGGCAAGCCATCTTGTCTAT	360
5247	Db		ACCAAGTTGGCCAGGTAGCGCTTTTCTCGGCGAACCATCGGCAAGCCATCTTGTCTAT	5306
361	Qy		GGCGATGTGTGATGCGTTCGACGAAAGCGGCATGCCCGACCTTTCGACCTCGACCGC	420
5307	Db		GGCGATGTGTGATGCGTTCGACGAAAGCGGCATGCCCGACCTTTCGACCTCGACCGC	5366
421	Qy		CTCCTATTGAGACGAATTTGTGCGACCAATCGATCTTTTACGCGCGGTGAGCTTTTCGAC	480
5367	Db		CTCCTATTGAGACGAATTTGTGCGACCAATCGATCTTTTACGCGCGGTGAGCTTTTCGAC	5426
481	Qy		GGCATCGGCGCTTACAACCTCGCTACCGAGTCTGGGCGGACTGGGACTTTCAATATTTCG	540
5427	Db		GGCATCGGCGCTTACAACCTCGCTACCGAGTCTGGGCGGACTGGGACTTTCAATATTTCG	5486
541	Qy		TGCTTCTCCAAACCGCGGCGTGATTACCGGTACATGGACGTCGTGATTTCCGAATACAAC	600
5487	Db		TGCTTCTCCAAACCGCGGCGTGATTACCGGTACATGGACGTCGTGATTTCCGAATACAAC	5546
601	Qy		GACATGACCGGCTTCAGATGAGGCGAGGAGCTGATAAGAGTTTCAGAAAACGGCTGCCA	660
5547	Db		GACATGACCGGCTTCAGATGAGGCGAGGAGCTGATAAGAGTTTCAGAAAACGGCTGCCA	5606
661	Qy		ATGTACTTCTGGGTTGACGGGTGGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAG	720
5607	Db		ATGTACTTCTGGGTTGACGGGTGGAGACTTGCAGGCGCATGCTGGCGTTTTTGAAGAG	5666
721	Qy		AAGGAGAAATCCCGCTCTGGCTTCGCTACGCGGTTGATAAGGGTTAAGGCGCTCTCCAAA	780
5667	Db		AAGGAGAAATCCCGCTCTGGCTTCGCTACGCGGTTGATAAGGGTTAAGGCGCTCTCCAAA	5726
781	Qy		GAAACGAGCGCAGAACCGTAG	801
5727	Db		GAAACGAGCGCAGAACCGTAG	5747

RESULT 5

```

US-10-282-122A-26361
: Sequence 26361, Application US/10282122A
: Publication No. US20040029129A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
:

```

QY 541 TGCTTCTCAACCCGCGCTGATTAACCCGCTACATGAGCGTGGTGAATTTCCGAATACAAAC 600
|||||
Db 541 TGCTTTTCCAAACCCAGCGCTCGTACCCGCTACATGACGTGGTGGTTCGAAAGCTACAAAC 600
|||||
QY 601 GACATGACCGGCTTACGATGAGCGGAGGAGTGTATTAAGAGTTTCAGAAACGGCTGCCA 660
|||||
Db 601 GAATTCGGCGGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTTTGAAGCGGCTGCCG 657
|||||

661 ATGTAC 666

658 ATGTCC 663

RESULT 6

US-10-282-122A-28624

; Sequence 28624, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangshu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining prior application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 28624

; LENGTH: 828

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-28624

Query Match 43.2%; Score 346; DB 17; Length 828;

Best Local Similarity 71.0%; Pred. No. 6.8e-104;

Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 1 ATGACTGCCCGAGTGTTCGATAATATCCCTACCTTCAATGCGAGCGGTGACGCTGCAA 60

|||||

Db 58 ATGGCGGACCAATGTTTTCGATCATCCCCACCTTGAACGTGGCTGGGTATTGGCT 117

|||||

QY 61 GCCTGCTCGGAAGCATCTCGGGCAGACCTACCGGGAAGTGGAGTGGTCTTGTGCGAC 120

|||||

Db 118 GCCTGCTCGAAGCATCGCCCGTCAGACCTGCGGTGAGCTTCGAGCTGGTACTGCTGAC 177
QY 121 GCGGCTTCGACCGATCGGACCTCGACATCGGAACAGTTTCGCGCCCGGAACCTCGGCTCG 180
|||||
Db 178 GCGGCTTCGACCGAAGCAACCTCGACATCGCAACATTTTCGCCCCCAACCTTCGGCGAG 237
|||||
QY 181 CGACTGGTTCGTTACAGCGGCGCCGATGATGGCCCTTACGAGCGCATGAACCGCGGCGTC 240
|||||
Db 238 CGGTTGATCATTCATCGCGACACCGACGAGGCGCTTACGACGCCATGAACCGCGCGGTG 297
|||||
QY 241 GCGGTAGCCACAGGCGAATGGTACTTTTATAGCGCGGCGAGCACACCTCTTACCAACA 300
|||||
Db 298 GACCTGGCCACCGGAACGTGGTGTCTCTTCTGGGCGCGGAGCAGAGCTGTACGAGGT 357
|||||
QY 301 ACCAGTTGGCCAGGTAGCGGCTTTTCTCGCGCAACCATGCGGCAAGCATCTTGTCTAT 360
|||||
Db 358 GACACCTTGGCGGGGTGCGGCTTCAATGGGGAACACAGAGCCAGCGATCTGATATAT 417
|||||
QY 361 GCGGATGTTGATGCGTTTCGACGAAAGCGGCAATGCGGACCTTTTCGACCTCGACCGC 420
|||||
Db 418 GCGGACGTGATCATGCGCTCAACCAATTTTCGCTGGGCTGGGCGCTTCGACCTCGACCGT 477
|||||
QY 421 CTCCTATTGAGACGAATTTGTGCCAACCAATCGATCTTTTACCGCGGTGAGCTTTTCGAC 480
|||||
Db 478 CTGTTGTTCAAGCGCAACATCTGCCATCAGGCGATCTTCTACCGCGCGGACTCTTTCGGC 537
|||||
QY 481 GGCATCGGCGCTTACAACTCGCTACCGAGTCTGGGCGGAGCTGGGAGCTTCAATATTCG 540
|||||
Db 538 ACCATCGGCTTCAACACTCGCTACCGGTCTCGGCTGCGGAGCTTCAATATTCG 597
|||||
QY 541 TGCTTCTCAACCCGCGCTGATTAACCGCTACATGAGCGTGGTGAATTTTCGAAATACAA 600
|||||
Db 598 TGCCTTTCACACCCAGCGCTCGTCACCGCTACATGACAGCTGGTGGTTCGAAAGCTACA 657
|||||
QY 601 GACATGACCGGCTTCAGCATGAGCGGAGTGTATTAAGAGTTTCAGAAACCGCTGCCA 660
|||||
Db 658 GAATTCGGCGGGCTCAGCAATACGATCG---TCGACAAGGAGTTTTTTGAAGCGGCTCG 714
|||||
QY 661 ATGTAC 666
|||||
Db 715 ATGTCC 720

RESULT 7

US-10-805-311-38

; Sequence 38, Application US/10805311

; Publication No. US20040260078A1

; GENERAL INFORMATION:

; APPLICANT: Hermon-Taylor, John

; APPLICANT: Doran, Tim

; APPLICANT: Millar, Douglas

; APPLICANT: Tizard, Mark

; APPLICANT: Loughlin, Mark

; APPLICANT: Sumar, Nazira

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC

; TITLE OF INVENTION: MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND

; TITLE OF INVENTION: TARGETS FOR CHEMOTHERAPY

; FILE REFERENCE: 117-260

; CURRENT APPLICATION NUMBER: US/10/805,311

; CURRENT FILING DATE: 2004-03-22

; PRIOR APPLICATION NUMBER: US/09/705,911

; PRIOR FILING DATE: 2000-11-06

; PRIOR APPLICATION NUMBER: US/09/091,538

; PRIOR FILING DATE: 1998-09-16

; PRIOR APPLICATION NUMBER: PCT/GB96/03221

; PRIOR FILING DATE: 1996-12-23

; PRIOR APPLICATION NUMBER: GB 9526178.0

; PRIOR FILING DATE: 1995-12-21

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 38

; LENGTH: 828

; TYPE: DNA

Best Local Similarity 52.8%; Pred. No. 1.2e-10;
Matches 149; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

```
QY 9 GCCAGTGTTCGATAATTTATCCCTACCTTCAATCAGCGGTGACGCTGCAAGCCTGCCT 68
Db 6184421 GCCAGTGTTCAGTGTATCGTCCCGGTACAGGTTTACGGGTACCTGCAATGAGTGCCT 6184362

QY 69 CGGAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGTTCCTTGTGACGGGGTTC 128
Db 6184361 CGAATCCGTCGTGGAGCAGTCTTCCCGACCTCGAATGATCGCGGTGACGACTGCTC 6184302

QY 129 GACCATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTGGT 188
Db 6184301 ACCCGACGCTCGCGGCGCATCATCGACGAGTTTCGGGGCCCGTGACACCCGGGTACGGCC 6184242

QY 189 CQTTCACAGCGGCGCGATGATGGCCCTACGACGCCATGAACCGCGGTGCGGTAGC 248
Db 6184241 CQTTCACCTGCCACAGAACAGCGGTCTGGGCGCGGCGCGACAGCGGCTGCGCGAGGC 6184182

QY 249 CACAGCGGAATGGGTACTTTTTTATAGGCGCGGACGACACCTT 290
Db 6184181 CACCGGCGACTACCTGATCTTCTTGGACGCGGACGACAGCCT 6184140
```

RESULT 10

```
US-10-156-761-5077
; Sequence 5077, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5077
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2196)
US-10-156-761-5077
```

Query Match 8.4%; Score 67.4; DB 15; Length 2196;
Best Local Similarity 52.3%; Pred. No. 2.6e-11;
Matches 149; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
QY 7 GGGCCAGTGTTCGATAATTTATCCCTACCTTCAATGACGGGTGACGCTGCAAGCCTGC 66
Db 1 GTGCGCGGCTTCAGCATCATCGTCCCGCTTCTTCAAGGTGCAGGGGTATCTGCGGAGTGC 60

QY 67 CTCGGAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGACGGCGGT 126
Db 61 CTCGACTCGGTCTCGGGCAGTGTGACCGGATCTGGAAGTATCGCGTTCGACACTGC 120

QY 127 TCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTG 186
Db 121 TCGCGGACGGCTGCGGCGCCATCTTCGACGGGTACGCGGCCGCGGACCGCGGTACGC 180

QY 187 GTGCTTCAGCGGGCGCGCATGATGGCCCTTACGACGCGCATGAACCGGGGGTTCGGCGTA 246
Db 181 GTGCTGCACCTCGCGGAGAACGTGCGGCTTGGGCGCGGACGCAACGCGGGAATGCGGCAC 240
```

```
QY 247 GCCACAGCGAATGGGTACTTTTTTATAGCGCCGCGACGACACCTC 291
Db 241 GCCACCGGGGACTACCTTCTTCTCTCGACAGCGAGACACCTC 285
```

RESULT 11

```
US-10-282-122A-30023
; Sequence 30023, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30023
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30023
```

Query Match 7.5%; Score 60; DB 17; Length 900;
Best Local Similarity 50.0%; Pred. No. 5.3e-09;
Matches 209; Conservative 0; Mismatches 200; Indels 9; Gaps 2;

```
QY 7 GCGCCAGTGTTCGATAATTTATCCCTACCTTCAATGACGGGTGACGCTGCAAGCCTGC 66
Db 46 GCGCGGTGTATCGTCTAGCGCCCTTCAACGGGAAAGTACTTGAAGAAGCC 105

QY 67 CTCGGAAGCATCGTCGGGCGAGACCTACCGGGAAGTGGAGTGGTCTTGTGACGGCGGT 126
Db 106 CTGCGGAGCATCTACGAGCAGGACTACCCGAATTTTGAAGTGATCATCTGTCGACGCGT 165

QY 127 TCGACCGATCGGACCTCGACATCGGAACAGTTTCCGCCCGGAACCTCGGCTCGGACTG 186
Db 166 TCCACCGAACACAGCTACGCCATCTGGAGCAGTTGAGAAAGTCCACGGTTTCCAGC-- 223

QY 187 GTCGTTACAGCGGGCGCGATGATGGCCCTTACGACGCGCATGAACCGCGGCGTTCGGCGTA 246
```


Search completed: March 6, 2005, 23:25:32
Job time : 617 secs
